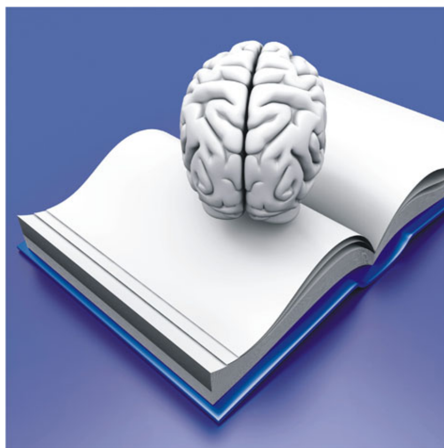
Psihologie

Florin A. Sava

PSIHOLOGIA VALIDATĂ ŞTIINŢIFIC

Ghid practic de cercetare în psihologie



Collegium

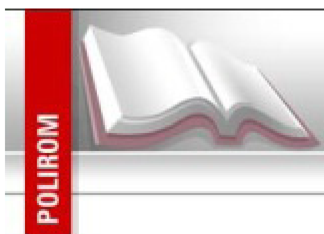
POLIROM

Florin A. Sava

PSIHOLOGIA
VALIDATĂ ŞTIINŢIFIC
Ghid practic de cercetare în psihologie

Prefață de Daniel David

POLIROM
2013



www.polirom.ro
Editura POLIROM
Iaşi, B-dul Carol I nr. 4; P.O. BOX 266, 700506
Bucureşti, Splaiul Unirii nr. 6, bl. B3A, sc. 1, et. 1;
sector 4, 040031, O.P. 53, C.P. 15-728

ISBN ePub: 978-973-46-3272-5
ISBN PDF: 978-973-46-3273-2
ISBN print: 978-973-46-2238-2

Coperta: Laurenţiu Midvichi
Pe copertă: © Spectral-Design/Shutterstock.com

FLORIN A. SAVA este reprezentantul domeniilor Psihologie şi Psihoterapie în cadrul Comisiei de Ştiinţe Sociale a Consiliului Naţional al Cercetării Ştiinţifice, directorul departamentului de Psihologie din cadrul Universităţii de Vest din Timişoara şi coordonatorul Laboratorului de Cogniţii Sociale şi Evaluare a Personalităţii (http://socpers.psihologietm.ro). Printre preocupările sale ştiinţifice se regăsesc evaluarea personalităţii şi a inteligenţei, precum şi măsurătorile şi intervenţiile implicite validate ştiinţific. Interesul pentru cercetare a fost completat, în plan naţional, cu activitatea de diseminare a cunoştinţelor din domeniul metodologiei cercetării şi al analizei datelor. De acelaşi autor: *Analiza datelor în cercetarea psihologică* (2004, 2011), *Metode implicite de investigare a personalităţii* (coord., Polirom, 2011).

Familiei mele, Laura, Rareş şi Sonia, pentru felul lor molipsitor de a se bucura în fiecare zi

Prefaţă

În domeniul psihologiei din România se publică foarte multe cărţi. Puţine sunt însă cele care contează cu adevărat, devenind repere la nivel naţional. Aceasta este o carte care contează şi care cred că va deveni un reper pentru psihologia românească.

Şcoala clujeană de psihologie este cea care a promovat în psihologia din ţară, în mod explicit şi programatic, metoda ştiinţifică experimentală. Într-adevăr, primul Institut de Psihologie Experimentală din România a fost înfiinţat în 1921 la Cluj de către profesorul Florian Ştefănescu-Goangă, fost student al lui Wilhelm Wundt, care este considerat la nivel internaţional fondatorul psihologiei. Autorul acestei lucrări, conferenţiarul Florin Alin Sava, este, prin studiile la nivel de licenţă şi masterat, un produs al şcolii de psihologie din Timişoara, iar prin studiile la nivel de doctorat şi postdoctorat, un produs al şcolii clujene de psihologie. Aşadar, substratul experimental şi ştiinţific riguros este fundamentul activităţilor academice şi de cercetare performante ale colegului Florin Alin Sava.

Şcoala clujeană de psihologie a generat după revoluţia din 1989 (nu mai pomenesc aici lucrările monumentale dinainte de revoluţie) o serie de publicaţii fundamentale pentru psihologia experimentală din România. Astfel, fundamentele iniţiale au fost puse de o lucrare coordonată de profesorul Ioan Radu (Radu şi colab., 1993, *Metodologie psihologică şi analiza datelor*, Sincron, Cluj-Napoca). Ulterior, Florin Alin Sava publică o lucrare care completează excelent lucrarea profesorului Radu, aducând în atenţia cititorului metode şi tehnici noi şi/sau detaliind critic metodele tradiţionale (Sava, 2004/2011, *Analiza datelor în cercetarea psihologică. Metode statistice complementare*, ASCR, Cluj-Napoca). Am extins apoi aceste fundamente experimentale în domeniul clinic (David, 2006, *Metodologia cercetării clinice. Fundamente*, Polirom, Iaşi). Prin lucrarea de faţă, Florin Alin Sava completează excelent lucrările menţionate anterior prin extinderea metodologiei cercetării psihologice în contextul practicii psihologice validate ştiinţific (*evidence-based practice*), cu accent, fără a fi exclusiv, pe domeniul clinic.

Lucrarea colegului Florin Alin Sava este una complexă, foarte bine gândită şi structurată, acoperind aspecte teoretice (de exemplu, problematica pseudoştiinţei), metodologice (de exemplu, metaanaliza, studiul clinic controlat) şi practice (de exemplu, raportarea rezultatelor în articole), foarte importante pentru psihologia modernă. Sigur, nu este o lucrare uşor de parcurs şi înţeles – fiind totuşi o lucrare ştiinţifică care abordează aspecte dificile –, dar este uşor de citit, fiind bine scrisă. În plus, lucrarea provoacă şi stimulează, devenind treptat un manual personal de *know-how*, la care ne putem raporta mereu în cercetări şi/sau publicarea articolelor ştiinţifice.

Aşa cum am afirmat anterior, cred sincer că această lucrare va deveni un reper în familia lucrărilor de metodologia cercetării în psihologie publicate în ţară, fiind necesară şi utilă unui public divers. De aceea o recomand cu toată încrederea nu doar studenţilor, ci oricărui cercetător serios din psihologie, în particular, şi din ştiinţele socioumane, în general.

Prof. univ. dr. Daniel David
Universitatea „Babeş-Bolyai”, Cluj-Napoca, România (profesor)
Mount Sinai School of Medicine, New York, SUA (profesor asociat)

Cuvânt înainte

Pentru cei mai mulţi dintre cei care ajung să răsfoiască sau să citească acest volum, psihologia reprezintă un domeniu fascinant şi complex, ce oferă o perspectivă aparte asupra realităţii înconjurătoare şi care, de ce nu, reprezintă un mijloc de trai prezent sau viitor. În interiorul psihologiei avem libertatea de a alege anumite perspective teoretice cărora le devenim loiali, în care credem şi pe care ajungem, la rândul nostru, să le promovăm către ceilalţi, fie aflaţi în interiorul breslei (studenţi, practicieni sau oameni din sfera academică), fie din afara ei (pasionaţii de psihologie şi beneficiarii serviciilor psihologice). Îndrăgim aceste modele şi explicaţii psihologice pentru că ni se potrivesc, pentru că rezonăm afectiv cu ele şi ajungem să le îmbrăţişăm, fără a căuta dovezi solide sau raţionamente logice, atent construite, pentru a le accepta.

Sloganul sub care se desfăşoară întreg acest scenariu este presărat cu intenţii bune, consumatorii de psihologie fiind invitaţi să îşi aleagă abordarea care li se potriveşte, pentru că este suficient loc pentru toate perspectivele teoretice în psihologie. Şi, într-adevăr, această diversitate productivă a permis psihologiei să evolueze şi să îşi extindă influenţa dincolo de graniţele sale clasice, dincolo de disputele tradiţionale dintre psihanaliză, behaviorism, abordarea umanistă sau cognitivism. Astăzi, din ce în ce mai mulţi psihologi lucrează în neuroştiinţele cognitive – pentru a stabili legătura dintre procesele cognitive şi creier; în genetică – pentru a stabili legătura dintre anumite structuri genetice şi comportament; în medicină – pentru a stabili legătura dintre anumite constructe psihologice

şi riscul îmbolnăvirii etc. Psihologia şi-a extins influenţa nu doar în zona ştiinţifică, ci şi în aceea a spiritualităţii. De exemplu, întreg domeniul New Age este saturat în explicaţii de natură psihologică. Parafrazând o expresie comună, aş spune că sub umbrela psihologiei a fost loc pentru toată lumea.

Filosofia acestei cărţi pune la îndoială o asemenea perspectivă liberală. Lucrarea este înrădăcinată în perspectiva abordării validate ştiinţific (*evidence-based approach*), iniţiată în interiorul ştiinţelor medicale. Această perspectivă este una ateoretică, prin urmare nu veţi găsi alte paradigme de explicare a comportamentului uman decât cele deja consacrate. Ceea ce face abordarea validată ştiinţific este să analizeze critic modelele teoretice existente în psihologie, utilizând aparatul ştiinţific. În termenii unei analogii, dacă presa este câinele de pază al democraţiei, am putea spune că abordarea validată ştiinţific este gardianul oricărei ştiinţe empirice. În absenţa vocii (perspectivei) abordării validate ştiinţific, psihologia este ameninţată să fie parazitată de abordări pseudoştiinţifice dăunătoare nu numai pentru ea ca ştiinţă, ci şi pentru beneficiarii serviciilor psihologice. În primele două capitole ale cărţii vom insista asupra educării publicului pentru a diferenţia între ştiinţă şi pseudoştiinţă în psihologie şi asupra prezentării premiselor şi paşilor de lucru în practica psihologică, astfel încât serviciile psihologice oferite să fie unele validate ştiinţific.

Astfel, la fel ca în alte ştiinţe preocupate de optimizarea comportamentului uman şi de rezolvarea unor probleme care pot afecta sănătatea şi calitatea vieţii unui individ, şi în psihologie avem nevoie: (a) să ştim că intervenţia oferită pentru problema constatată este una eficientă (în urma unui demers ştiinţific); (b) să ştim de ce intervenţia oferită este una eficientă, pentru a-i potenţa efectul şi pentru a cunoaşte mai bine

mecanismele psihologice implicate (în urma unui demers ştiinţific); (c) să ştim, din respect pentru clienţii noştri, dar şi din constrângeri bugetare, că soluţia aleasă pentru problema lor are cel mai bun raport cost-eficienţă (desigur, tot printr-un demers ştiinţific). Abordarea validată ştiinţific are în vedere toate cele trei puncte menţionate, dar mai cu seamă pe primele două.

Celelalte şase capitole sunt rezervate subiectului principal al lucrării de faţă, adică descrierii demersului de cercetare necesar pentru a afirma despre un serviciu psihologic oferit că este unul validat ştiinţific (adică un serviciu eficient, fundamentat pe un model teoretic valid şi cu un bun raport cost-eficienţă). Astfel, capitolul 3, unul dintre cele mai importante capitole ale lucrării, prezintă o grilă de lectură a studiilor empirice din psihologie, prin componentele critice necesare realizării şi/sau evaluării unui demers de cercetare. În cadrul capitolului am subliniat în numeroase rânduri importanţa dinamicii dintre obiectivele cercetării (ceea ce-şi propune studiul) şi validitatea cercetării (metodologia şi mijloacele de analiză a datelor utilizate pentru a răspunde adecvat la obiectivul studiului).

Demersul ştiinţific este, apoi, detaliat în patru capitole succesive (4-7), fiecare prezentând tipurile principale de studii întâlnite: cele experimentale şi cvasiexperimentale (capitolul 4), cele non-experimentale (capitolul 5), studiile de tip cost-eficienţă (capitolul 6) şi cele metaanalitice (capitolul 7). Lucrarea se încheie cu un capitol destinat redactării articolelor ştiinţifice (capitolul 8), scris din dorinţa de a ajuta studenţii, masteranzii şi doctoranzii să înţeleagă standardele de publicare din domeniul psihologiei, să-şi întărească simţul critic atunci când lecturează rezultatele unor studii şi, de ce nu, să-i ajute să scrie mai bine lucrările ştiinţifice pe care le au de redactat. Toate aceste cinci capitole

sunt scrise într-o manieră practică, oferind cititorilor o serie de sugestii şi recomandări despre cum să deruleze astfel de studii şi cum să reflecteze critic asupra rezultatelor obţinute, constituind, astfel, un veritabil manual de metodologie a cercetării. Utilitatea conţinuturilor prezentate depăşeşte cadrul discuţiei despre psihologia validată ştiinţific, informaţiile fiind relevante pentru a înţelege demersul de cercetare din întreaga zonă a ştiinţelor comportamentale. Prin complexitatea unor analize prezentate, lucrarea răspunde mai bine nevoilor specialiştilor în psihologie decât amatorilor de psihologie. Specialiştii în psihologie vor găsi în acest volum acele informaţii care să-i ajute să facă pasul de la textele introductive de metodologia cercetării la cerinţele necesare pentru a analiza critic rezultatele studiilor lecturate şi de a efectua cercetări competitive la nivel internaţional. Amatorii de psihologie vor înţelege mai bine eforturile psihologiei ca ştiinţă, astfel încât să ajungă să distingă între serviciile psihologice bine fundamentate şi abordările ezoterice. Acest lucru îi va ajuta să devină consumatori inteligenţi de psihologie şi să preţuiască actul psihologic fundamentat ştiinţific.

În final, aş dori să mulţumesc celor care au contribuit la efortul de a duce la bun sfârşit acest proiect. Aş începe cu cei care au fost primii mei cititori şi care m-au ajutat să calibrez complexitatea conţinuturilor prezentate spre beneficiul masteranzilor şi al doctoranzilor. Este vorba de Otilia Matei (masterandă) şi de colaboratorii mei cei mai apropiaţi, doctoranzii Andrei Rusu şi Silvia Rusu. Apoi i-aş aminti pe Dr. Laurenţiu Maricuţoiu şi Dr. Sebastian Pintea, a căror consultanţă tehnică oferită a fost extrem de folositoare. Nu în ultimul rând, îi mulţumesc profesorului universitar clujean Daniel David pentru feedbackul său consistent şi pertinent legat de conţinutul unor capitole, dar mai ales pentru efortul susţinut pe care-l depune în promovarea psihologiei validate ştiinţific ca

reper prescriptiv în activitatea oricărui psiholog. Este un deziderat în care cred şi în folosul căruia sper să contribuie această carte, prin impactul ei în comunitatea psihologilor din România. Am lăsat la final mulţumirile pentru cei alături de care încep mereu o nouă zi, Sonia, Rareş şi Laura. Le mulţumesc pentru suportul afectiv şi pentru momentele de inspiraţie prilejuite.

Florin A. Sava

Timişoara, 2013

Partea întâi

Fundamentele cercetării ştiinţifice în psihologie

Capitolul 1
Abordarea ştiinţifică *vs* pseudoştiinţifică în psihologie

1.1. Expansiunea psihologiei în societate

Îndeosebi după cel de-al Doilea Război Mondial, psihologia a devenit un domeniu ştiinţific şi aplicativ aflat într-o creştere exponenţială, ce continuă inclusiv în prezent. O statistică realizată de Sexton şi Hogan reprezentativă pentru SUA (1992 *apud* Louw, 2002) plasează psihologia pe locul trei, ca ritm de creştere a numărului de studenţi, după informatică şi matematică. Această situaţie nu este însă una izolată, ea fiind prezentă în foarte multe ţări europene, precum şi în statele cu o economie bine dezvoltată cum ar fi Australia, Canada etc. Bunăoară, numai în Australia, contribuţia serviciilor psihologice la produsul intern brut a crescut în zece ani cu 500% faţă de 1991, iar numărul studenţilor care urmează studii de master sau doctorat din domeniul psihologiei aproape s-a dublat în aceeaşi perioadă (Patrick, 2005). Aceeaşi situaţie, ce reflectă un interes deosebit pentru psihologie, se regăseşte şi în societatea românească. De pildă, datele de mai jos descriu o situaţie particulară, cea a admiterii la Universitatea de Vest din Timişoara: în sesiunea din iulie 2011, psihologia a ocupat locul 3 ca număr de studenţi înscrişi în anul I, după ştiinţele economice şi drept. Serviciile psihologice nu au crescut însă numai

cantitativ, ci s-au şi diversificat, dezvoltând noi nişe de piaţă din perspectiva serviciilor psihologice oferite (Goodheart, 2010). Cu toate că majoritatea serviciilor de psihologie se regăsesc în sfera psihologiei clinice şi a consilierii psihologice (55% dintre psihologii din întreaga lume fiind psihologi clinicieni sau consilieri psihologici – vezi Sexton şi Hogan, 1992 *apud* Louw, 2002), în ultima vreme asistăm la o creştere a serviciilor psihologice în arii din ce în ce mai variate, precum domeniile industriei şi afacerilor, juridic, militar, sportiv, în publicitate şi politică etc.[1].

O posibilă explicaţie, de sorginte filosofică, pentru expansiunea psihologiei constă în nivelul ridicat de insecuritate pe care îl resimte o persoană în încercarea sa de a se adapta la provocările lumii contemporane (Savage, 2009), psihologia fiind domeniul care oferă indivizilor afectaţi de aceste provocări, modalităţi adecvate de coping, potrivite unei societăţi de tip secularizat.

O altă explicaţie de tip filosofic pune expansiunea psihologiei pe seama apariţiei societăţii capitaliste postindustriale de tip occidental, care se caracterizează printr-un sistem liberal avansat, ce conferă cetăţenilor săi o serie de drepturi şi libertăţi. Valori precum individualismul sau libertatea de a decide ocupă un rol principal în discursul public, iar grija primordială a guvernanţilor este de a oferi cât mai multor indivizi o viaţă în care sănătatea, fericirea şi bunăstarea personală sunt scopuri de atins (Jansz, 2004, Louw, 2002). Din acest punct de vedere, psihologia este domeniul ştiinţific care poate contribui eficient la bunăstarea individuală. Discursurile din mass-media despre nevoile individuale ale fiecăruia şi despre dreptul fiecăruia dintre noi de a-şi satisface aceste nevoi („pentru că merit!”), dublate de industria de tip *self-help* care a dobândit un caracter de masă („pentru a evolua emoţional şi spiritual!”) au condus la

dezvoltarea abordării de tip psihologic în societatea contemporană. Această explozie în mass-media, dublată de preocuparea cinematografiei de pretutindeni pentru ecranizarea unor scenarii cu încărcătură psihologică, au contribuit la constituirea unei societăţii psihologizate (Jansz, 2004), fapt ce i-a determinat pe unii autori precum Gross (1979, p. 3 *apud* Louw, 2002) să afirme: „Astăzi, psihologia este artă, ştiinţă, terapie, religie, cod moral, stil de viaţă, filosofie şi cult… miezul societăţii contemporane”.

În sfârşit, o a treia explicaţie posibilă pentru expansiunea psihologiei ca ştiinţă, îşi are originea în modificarea implicării statului în menţinerea şi promovarea sănătăţii cetăţenilor săi, prin reconceptualizarea modului în care se face asistenţa socială şi asistenţa medicală acordată populaţiei. Acest lucru a condus nu numai la dezvoltarea serviciilor de psihologie clinică şi consiliere psihologică, ci şi la pătrunderea cunoştinţelor de psihologie în pregătirea profesională a altor ocupaţii din sfera asistenţei medicale şi sociale. Pe lângă beneficiile aduse de includerea serviciilor psihologice în practica asistenţei medicale/sociale, care au condus la creşterea cererii de servicii psihologice pe piaţă, această practică a condus, adeseori, la percepţia greşită din partea oamenilor de rând că psihologia se ocupă doar de oameni aflaţi în situaţii de suferinţă emoţională ca urmare a instalării unor tulburări psihopatologice sau a problemelor familiale/profesionale ce necesită consiliere (Levant, 2003).

1.2. Costurile expansiunii sociale asupra psihologiei ca ştiinţă

Beneficiile expansiunii psihologiei în societatea actuală sunt însoţite, însă, de o serie de costuri pe care psihologia, ca domeniu ştiinţific le plăteşte adesea. Pe scurt ele s-ar putea încadra în patru mari categorii: (a) superficialitatea discursului psihologic public; (b) distorsionarea conţinutului public referitor la psihologie; (c) asocierea în mod eronat a psihologiei cu spiritualitatea; (d) dezvoltarea unei industrii alternative bazată pe pseudoştiinţe.

1.2.1. Superficialitatea discursului psihologic public

Într-o bună măsură, publicul a prins gustul discuţiilor şi interpretărilor cu iz psihologic. În cercurile de intelectuali şi filosofi, cât şi în domeniul ştiinţelor umaniste, influenţa lui Freud şi a constructelor psihanalitice a devenit una relevantă pentru evoluţia domeniului. Publicul larg, prin intermediul emisiunilor TV sau a articolelor din presă, a incorporat în vocabularul său expresii de genul „criză de identitate”, „reprimarea nevoilor”, „dezvoltare personală”, „stimă de sine” etc. Dezbaterile de acest gen au dezvoltat publicului senzaţia că în psihologie, ca în politică, oricine îşi poate exprima o părere competentă, că discursul psihologic este unul facil, de bun-simţ. Din nefericire, la această impresie au contribuit inclusiv specialiştii prezenţi în mass-media, psihologi sau psihoterapeuţi cu notorietate, dar fără contribuţii ştiinţifice relevante în domeniu. În interpretarea diferitelor evenimente, ei apelează frecvent, fie la explicaţii tipice simţului comun, lipsite de specificitatea unei adevărate relaţii de tip cauză-efect (de exemplu, „problemele adolescenţilor provin aproape întotdeauna din criza lor de identitate şi din conflictul existent între generaţii”), fie, mai grav, promovează o serie de credinţe populare, dar

false din punct de vedere ştiinţific (de exemplu, „majoritatea oamenilor folosesc doar 10% din capacitatea creierului”). Pentru o listă de mituri (credinţe populare) fără fundamentare ştiinţifică puteţi consulta traducerea lucrării lui Lilienfeld, Lynn, Ruscio şi Beyerstein (2010).

1.2.2. Distorsionarea conţinutului public referitor la psihologie

În prefaţa aceleiaşi lucrări (Lilienfeld *et al.*, 2010, pp. 13-14) se afirma că „psihologia este peste tot în jurul nostru… în fiecare zi, ştirile, programele de televiziune, filmele şi internetul ne bombardează cu tot felul de presupoziţii în ceea ce priveşte o mulţime de subiecte psihologice… în lumea grăbită a supraîncărcării informaţionale, *dezinformarea* referitoare la psihologie este cel puţin la fel de răspândită ca informaţia corectă”.

Acest lucru poate fi observat direct la o parte însemnată dintre studenţii anului I, ciclul licenţă, care au aşteptări nerealiste despre ce vor învăţa la specializarea psihologie. În mentalul lor se regăsesc mai degrabă topici apropiate de ideea senzaţionalului (de exemplu, tulburările mentale şi abuzul sexual asupra copiilor) sau topici legate de domeniul parapsihologiei (de exemplu, experienţele în afara corpului, visele premonitorii etc.), decât topici specifice unei abordări ştiinţifice a psihologiei precum neuropsihologia sau învăţarea bazată pe condiţionare operantă.

Distorsionarea conţinutului public al psihologiei are loc şi la un alt nivel, cel al diferenţelor între agenda publică (psihologia populară) şi agenda psihologiei ştiinţifice. În acest sens, Stanovich (2010) remarca foarte bine că există o discrepanţă majoră între temele de interes public de sorginte psihologică şi temele actuale din cercetarea psihologică. De

pildă, pentru public, cele mai cunoscute teorii psihologice sunt cele dezvoltate de Freud şi de susţinătorii săi. Acestea constituie cel mai adesea reprezentarea în mentalul colectiv a domeniului psihologiei. În abordarea ştiinţifică, însă, conceptele lui Freud constituie adesea informaţii complementare prezentate în orice manual introductiv de psihologie, reprezentând doar 1-2% din conţinutul psihologic prezentat. Mai mult, cel mai adesea aceste informaţii sunt prezentate într-un context critic, subliniind limitele teoriei psihanalitice.

1.2.3. Asocierea eronată a psihologiei cu spiritualitatea

Poate e greu de crezut, dar cu aproximativ 100 de ani în urmă prestigioasa universitate britanică Cambridge a refuzat înfiinţarea unui laborator de psihofizică deoarece acesta ar „insulta religia, prin faptul de a măsura sufletul omului prin intermediul unui set de scale” (Hearst, 1979 *apud* Stanovich, 2010, p. 17). Această tendinţă de a intersecta (asocia) psihologia cu problemele de spiritualitate a rămas constantă în timp, iar în urma secularizării societăţii actuale şi a dezvoltării unor direcţii de studiu precum psihologia transpersonală sau psihologia umanistă s-a obţinut chiar un discurs ce îmbină concepte psihologice şi spirituale, atractiv pentru o bună parte a publicului. Acest lucru este valabil îndeosebi pentru adepţii unei viziuni holiste (Kendler, 1999), care au tendinţa de a căuta explicaţii de sorginte spirituală despre funcţionarea organismului uman, tinzând să caute mereu o explicaţie de genul „întregul este mai mult decât suma părţilor” şi să folosească eventualele dovezi ştiinţifice colaterale drept punct central de sprijin al ipotezei lor că există ceva dincolo de lumea materială. Amalgamul de dovezi ştiinţifice, de filosofie şi de idei

religioase (spirituale) poate fi extrem de atrăgător pentru aceste persoane aflate în căutarea fericirii şi confortului spiritual, dar dăunător pentru psihologie ca domeniu ştiinţific. Această stare de fapt a condus la o explozie a literaturii de tip *self-help*, cele mai multe dintre aceste cărţi sunt făcute să se vândă, nu să ajute, fiind arareori fundamentate ştiinţific (Kaminer, 1992). Aceeaşi preocupare pentru spiritualitate a condus la integrarea unor metode şi tehnici psihologice într-un domeniu mai larg, cel al medicinei alternative.

1.2.4. Proliferarea unei industrii alternative bazată pe metode pseudoştiinţifice

Înclinaţia minţii umane spre a accepta drept valide anumite credinţe nefundamentate ştiinţific este extrem de ridicată. În SUA există de 20 de ori mai mulţi astrologi decât astronomi, iar 50% dintre adulţii americani cred în posibilitatea de a comunica cu cei decedaţi, în vindecarea prin credinţă sau în astrologie (Lilienfeld, Lohr şi Morier, 2001). Situaţia este similară şi în alte părţi, inclusiv în România, unde un sondaj recent a reflectat că 40% dintre români cred că astrologia le influenţează viaţă, iar aproape 50% (*sic*!) cred în continuare că Soarele se învârte în jurul Pământului, în timp ce o treime sunt absolut convinşi că există oameni posedaţi de diavol (Vlăsceanu, Duşa şi Rughiniş, 2010). Un episod elocvent, de notorietate, întâlnit la noi este cel din campania prezidenţială de la finalul anului 2009, când unul dintre candidaţi l-a acuzat pe celălalt că utilizează proprietăţile „flăcării violete” pentru a fi protejat energetic, demers prezentat pe larg în mass-media, dar sancţionat prin intermediul unei scrisori deschise cu iz pamfletar (David, 2010).

Proliferarea pseudoştiinţelor pe un asemenea tărâm mental devine una evidentă, iar psihologia, ca domeniu ştiinţific a fost, la rândul său, afectată. Trebuie precizat de la început că pseudoştiinţa nu este determinată de obiectul studiat, ci de modul de raportare la metodele ştiinţifice (Lilienfeld, Lynn şi Lohr, 2004). Prin pseudoştiinţă (însemnând *ştiinţa falsă*) se înţelege un demers aparent ştiinţific, dar care îşi apără descoperirile deviind de la metodele ştiinţifice, mai ales atunci când rezultatele empirice obţinute neagă concluziile pe care le susţin. În termeni plastici, pseudoştiinţele joacă rolul unor impostori, pretinzând că joacă după regulile ştiinţei, pe care totuşi nu le respectă. În urma acestei abordări, oamenii cu un nivel mai scăzut de pregătire în cultura şi metodologia ştiinţifică pot ajunge să considere asemenea demersuri ca fiind valide, deşi nu este cazul (Lilienfeld şi Landfield, 2008). Pseudoştiinţele pot lua atât forme grosiere de manifestare (cum ar fi cazul frenologiei), cât şi forme subtile, adesea sub forma unor tehnici variate de intervenţie în vederea optimizării comportamentului uman (neuroprogramarea lingvistică, terapia strigătului primar, tehnicile de autosugestie prin mesaje subliminale etc.).

1.3. Descriptori specifici ai pseudoştiinţelor

Distincţia dintre ştiinţe şi pseudoştiinţe se poate face pe seama unor indicatori. Cu cât o abordare particulară are mai multe caracteristici specifice pseudoştiinţei, cu atât mai uşor este de diferenţiat această abordare în sfera pseudoştiinţelor. Cu cât o abordare specifică are mai puţine caracteristici specifice pseudoştiinţei, cu atât este mai dificil de

diferenţiat în privinţa acelei abordări dacă este o abordare ştiinţifică sau pseudoştiinţifică (Lilienfeld *et al.*, 2004; Stanovich, 2010). Aceşti indicatori ar putea fi grupaţi în indicatori ai metodei (ce vizează preponderent aspecte ale metodologiei ştiinţifice, greu de decelat de publicul necunoscător), respectiv indicatori de stil (ce vizează preponderent aspecte ideologice, mai uşor de decelat de către public, chiar fără cunoştinţe avansate de metodologia cercetării).

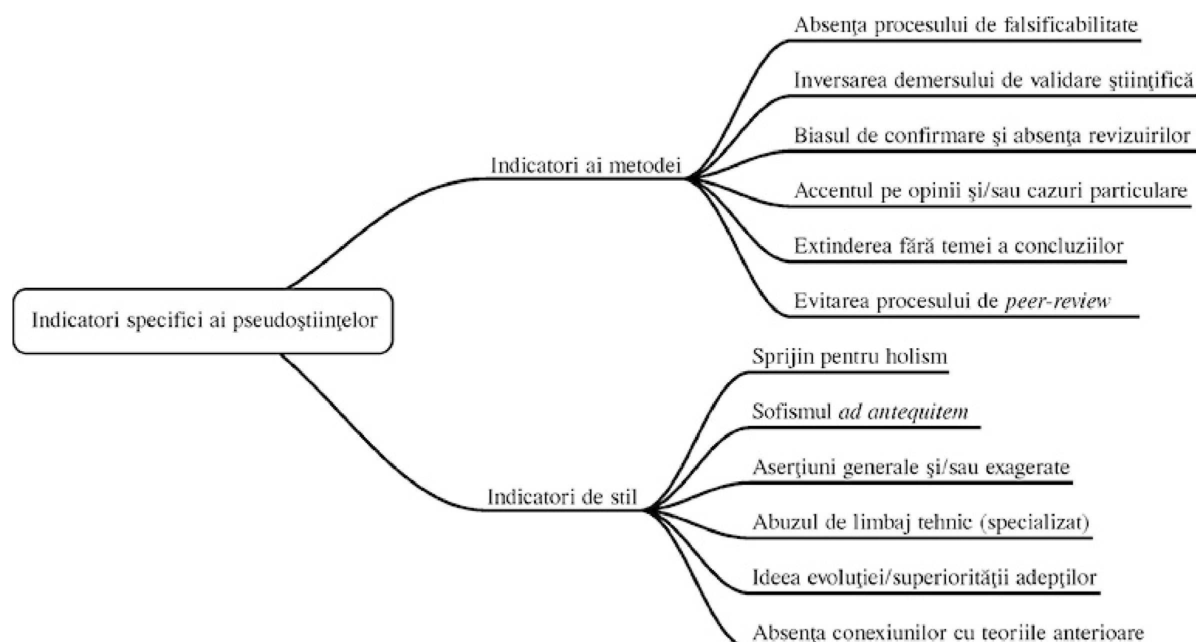


Figura 1.1. *Indicatori specifici ai pseudoştiinţelor*

Pentru a înţelege mai bine semnificaţia celor 12 indicatori amintiţi mai sus, vom descrie succint ce semnifică fiecare dintre ei.

1.3.1. Absenţa procesului de falsificabilitate

Conform lui Popper (1934/1981), la baza diferenţei dintre ştiinţă şi abordările neştiinţifice se află încercarea omului de ştiinţă de a strânge date empirice care să-i invalideze ipotezele avute. Demersul său este,

aşadar, unul indirect, asemenea procesului de testare a ipotezelor, în care rezultatele nu se contorizează în termeni precum „ipoteza cercetării se confirmă”, ci în aceia că „ipoteza nulă este respinsă, sprijinind astfel ipoteza cercetării”. Niciodată o teorie nu poate fi confirmată, ci doar sprijinită de datele empirice.

Pentru ca o teorie sau o abordare anume să poată fi falsificabilă este necesar (Stanovich, 2010): (a) să existe instrumente care să permită investigaţia ştiinţifică (de exemplu, expresia „Dumnezeu este creatorul universului” poate fi adevărată sau falsă, dar atât timp cât metodele ştiinţifice actuale nu permit verificarea afirmaţiei, nu se poate oferi un răspuns ştiinţific acceptabil deoarece lipsesc instrumente ştiinţifice care să permită testarea ei; (b) să fie suficient de specifică, încât să permită falsificabilitatea enunţurilor formulate (o teorie care poate explica orice sau care este prea vagă în formulare nu este considerată ca fiind ştiinţifică, deoarece nu se supune acestui criteriu al falsificabilităţii). De pildă, una dintre criticile aduse psihanalizei este aceea, că deşi teoria nu are capacitate predictivă specifică, cu alte cuvinte nu reuşeşte să anticipeze evoluţia unor comportamente, ea încearcă să explice orice *post factum*. Însă această capacitate explicativă *post factum* (de exemplu, „autismul este o expresie a unei comunicări deficitare, reci, dintre mamă-copil, a unui deficit de afecţiune resimţit de către copil ca venind din partea mamei”, o teorie psihanalitică populară în secolul XX, dar nesusţinută ştiinţific de date) poate face mai mult rău şi chiar stigmatiza anumite categorii de oameni. De aceea, o teorie ştiinţifică bună nu este una care încearcă să explice orice, într-o manieră raţionalizatoare, *post factum*, ci una care face predicţii specifice, anticipă evoluţia unor comportamente, iar aceste predicţii sunt susţinute de datele empirice colectate.

Caseta 1.1. Exemplu ipotetic de teorie pseudoştiinţifică

Un cercetător formulează teoria transferului comportamentului agresiv părinte-copil prin intermediul mecanismului inconştient de identificare cu agresorul. Conform acestei teorii, copiii preiau (învaţă) de la părinţii lor modul agresiv de a răspunde la situaţii frustrante, deoarece au observat că un asemenea comportament poate aduce beneficii autorilor actului agresiv (de exemplu, obţin ceea ce îşi doresc). Acest transfer se realizează într-o manieră inconştientă, în cele mai multe situaţii, cu excepţia cazului în care acest proces inconştient este blocat de un mecanism pulsional declanşat de instinctul de afiliere al copilului, care inhibă orice manifestare de tip agresiv orientată către semeni.

Acest gen de teorie poate explica *post factum* orice tip de scenariu, prin urmare nu permite testarea falsificabilităţii afirmaţiilor. Dacă copilul nu se comportă agresiv, explicaţia este că instinctul de afiliere a blocat transferul agresivităţii. Dacă copilul se comportă agresiv, explicaţia este o mărturie a transferului inconştient de identificare cu agresorul. Cu alte cuvinte, teoria poate fi utilizată pentru a explica orice, o caracteristică atractivă pentru consumatorii de psihologie cotidiană (pop), ce caută semnificaţii pentru originea comportamentului social, dar ineficientă din punct de vedere ştiinţific. Apoi, chiar dacă această teorie ar fi permis verificarea unor predicţii specifice (precum corelaţia dintre nivelul agresivităţii dintre părinţi şi copii), ar fi fost nevoie de mai multe dovezi care să sprijine specificitatea mecanismelor cauzale enunţate în teorie, precum transferul inconştient al comportamentului agresiv prin declanşarea mecanismului de apărare – identificarea cu agresorul; existenţa unui blocaj pulsional (*sic*!) determinat de constructul de tip umbrelă – instinctul de afiliere etc.

1.3.2. Inversarea demersului de validare ştiinţifică

În mod firesc îi revine cercetătorului care propune un anumit model sau o teorie să strângă dovezi în favoarea concepţiei sale. În cazul curentelor pseudoştiinţifice, acest aspect este inversat, solicitându-li-se scepticilor să ofere dovezi împotriva aserţiunilor făcute în teoria sau modelul respectiv. Mai concret, de exemplu, nu este datoria celor care contestă aserţiunea „efectele psihologice pozitive induse de practicarea meditaţiei sunt

cauzate de accesarea conştiinţei universale” să demonstreze scepticismul lor prin dovezi. Nici lipsa dovezilor, de ambele părţi, nu poate fi interpretată ca un indicator al susţinerii teoriei accesului la „conştiinţa universală” prin meditaţie, deoarece într-o abordare ştiinţifică este sarcina celui care propune o anumită teorie să ofere dovezi în sprijinul acesteia.

Caseta 1.2. *Sofismul de tip* ad ignoratium *utilizat frecvent pentru a inversa demersul ştiinţific*

În cei zece ani de practică psihoterapeutică folosind noua tehnică de terapie dinamică sinergetică nu s-a publicat niciun studiu care să demonstreze că tehnica ar fi una ineficientă. Prin urmare, această formă de terapie este o tehnică validă de intervenţie psihoterapeutică.

1.3.3. Biasul de confirmare şi absenţa revizuirilor

O diferenţă, sesizabilă pe termen lung, între teoriile ştiinţifice şi cele pseudoştiinţifice constă în acceptarea revizuirilor teoretice specifică abordărilor ştiinţifice. Ştiinţa a fost deschisă mereu la corecţii şi adaptări ale teoriilor stipulate în acord cu datele obţinute. În schimb, abordările pseudoştiinţifice au suferit puţine modificări de-a lungul timpului. De exemplu, astrologia, ce-şi propune să facă predicţii asupra evoluţiei unei persoane în funcţie de poziţionarea exactă a astrelor la momentul naşterii, a rămas aproape neschimbată de 4.000 de ani, cu toate că mişcarea axei Pământului în acest interval de timp a schimbat poziţia astrelor pe cer comparativ cu cea existentă cu patru milenii în urmă) (Hines, 2003 *apud* Lilienfeld şi Landfeld, 2008). Mai mult, pseudoştiinţele tind să se raporteze în mod subiectiv la rezultatele eventualelor studii efectuate, acceptându-le preferenţial doar pe acelea care le sunt convenabile, care

sunt în acord cu teoria stipulată. Dacă rezultatele susţin teoria enunţată, pseudoştiinţele folosesc aceste date ca mijloace pentru *advocacy*-ul şi marketingul teoriei respective. Dacă rezultatele nu susţin teoria enunţată, există tendinţa de a ignora (minimaliza) aceste rezultate şi de a găsi tot felul de argumente (scuze) printre limitele metodologice ale studiului pentru a susţine în continuare teoria enunţată. O asemenea tendinţă este observată şi printre abordările ştiinţifice, însă în acest caz procesul de evaluare colegială (*peer-review*) joacă rolul de a limita concluziile formulate, oricare ar fi acestea, ţinând seama de limitele metodologice pe care le are orice demers ştiinţific.

1.3.4. Accentul pe opinii şi pe cazuri particulare

În justiţie, opiniile experţilor au o valoare însemnată în decizia pe care o ia judecătorul unui caz. În ştiinţă, opiniile experţilor sunt irelevante dacă nu sunt susţinute de date empirice provenite din studii riguroase din punct de vedere metodologic. Stanovich (2010) remarca, cu umor, că în medicină au fost formulate opinii favorabile atât din partea unor experţi, cât şi a unor beneficiari cu privire la valoarea curativă a dinţilor de porc, a excrementelor de crocodil, a prafului de pe mumiile egiptene şi a multor alte remedii trăsnite. Nici „înţelepciunea populară” românească nu rămâne restantă la nivel imaginativ când este vorba de remedii curative naturiste pentru tot felul de probleme medicale sau psihologice. În toate aceste cazuri, tratamentele nu pot distinge, pe baza unei simple mărturii, fie ea televizată şi sinceră, dacă efectul produs de acestea este unul real sau unul determinat de efectul placebo. Am putea afirma că întreaga istorie a medicinei, înainte de secolul XX este doar o istorie a efectului placebo (Postman, 1986). Dar vindecarea nu este doar dependentă de

cauza sugerată sau de efectul placebo. Pot apărea alte situaţii precum existenţa unor variabile confundate, regresia către medie, remisia spontană, maturarea, evoluţia ciclică a tulburării etc. Toate acestea pot conduce la o evoluţie favorabilă a unei tulburări (deficienţe). Studiile de caz prezintă aceleaşi neajunsuri. Ele pot fi utile la nivel ştiinţific pentru generarea de ipoteze, dar nu şi pentru susţinerea (confirmarea) acestora. Opiniile experţilor, bazate sau nu pe anumite studii de caz, reprezintă primul pas într-un demers ştiinţific, ce trebuie completat (verificat) prin realizarea de studii empirice. Pseudoştiinţele ignoră acest pas esenţial, preferând utilizarea mărturiilor experţilor şi a datelor provenite din unele studii de caz, pe care le prezintă drept dovezi ştiinţifice de netăgăduit, despre eficienţa unei intervenţii anume.

1.3.5. Extinderea fără temei a concluziilor

Orice program educaţional de metodologia cercetării subliniază ideea că rezultatele ştiinţifice trebuie privite în termenii limitelor pe care le-a avut studiul respectiv. Ideea validităţii concluziilor extrase pe baza rezultatele obţinute într-o anumită cercetare este centrală pentru orice abordare ştiinţifică, de aceea ea a fost abordată şi dezvoltată pe larg în mai multe secţiuni ale acestei lucrări. Pseudoştiinţele nu sunt preocupate de acest aspect. Dimpotrivă, ele tind să nu diferenţieze, de pildă, între studii bazate pe surprinderea mecanismelor schimbării şi studii de tip *efficacy* sau *effectiveness* (vezi tabelul 4.4). De asemenea, pentru pseudoştiinţe are puţină relevanţă dacă studiul este unul experimental sau unul non-experimental. Prin urmare, nu ar fi exclus să constatăm că autorii unui simplu studiu comparativ non-experimental ce urmărea relaţia dintre muzica de relaxare şi nivelul anxietăţii de stare, să ajungă la concluzii

extravagante, tipice unui demers pseudoştiinţific, precum cel prezentat în caseta 1.3.

Caseta 1.3. *Exemplu de extindere neîntemeiată a concluziilor*

Diferenţele semnificative statistic înregistrate între cei care au urmat şi cei care nu au urmat şedinţe de meloterapie, indică un nivel mai scăzut al anxietăţii în cazul primului grup. Acest lucru îmbucurător este cauzat de fenomenul de rezonanţă ce apare între compoziţia vibratorie specială a muzicii şi mintea celui care o ascultă, inducând ascultătorului o stare psihică specială, de linişte şi armonie interioară. Muzica, ne pune aşadar, în rezonanţă cu energii benefice din macrocosmos, influenţând benefic atât starea fizică, cât şi pe cea psihică.

1.3.6. Evitarea procesului de peer-review

Publicul tinde să se încreadă mai mult într-o anumită teorie dacă aceasta a fost publicată în reviste de specialitate. Pentru aceasta, ştiinţa apelează la procesul de evaluare colegială, cunoscut sub denumirea generică din limba engleză de *peer-review*. Cu toate că nu este un mecanism perfect, acest proces este menit tocmai pentru a elimina de la publicare acele articole care au deficienţe metodologice majore şi/sau care prezintă concluzii eronate, ce nu sunt susţinute de datele empirice colectate. În cele mai multe cazuri, pseudoştiinţele ocolesc acest proces, promovându-şi produsele direct, prin mijloace specifice marketingului comercial. Pentru a-şi justifica această decizie, promotorii pseudoştiinţelor acuză revistele de specialitate că promovează ortodoxismul ştiinţific, că dau dovadă de conservatorism şi de rigiditate în politica editorială. Această poziţie de „victimă" poate fi augmentată de enumerarea unor cazuri celebre de erori comise prin procesul de *peer-review*, conform cărora conceptele prea inovative pentru timpul în care au fost dezvoltate sunt

privite cu circumspecţie de către comunitatea ştiinţifică, la fel cum s-a întâmplat şi în cazul teoriei relativităţii a lui Einstein.

1.3.7. Sprijin pentru holism

Un mecanism de autoapărare prezent în cadrul multor pseudoştiinţe este apelul la complexitatea sistemului psihologic. Atunci când teoriile sau enunţurile făcute nu corespund cu datele empirice colectate, auzim mereu explicaţii de genul: lucrurile sunt mult mai complexe şi nu trebuie privite izolat (Lilienfeld *et al.*, 2004). Fie că este vorba de interpretarea unui element dintr-un test proiectiv (care mereu trebuie privit în mod contextualizat, coroborat cu informaţiile din anamneza pacientului, pentru a confirma presupoziţiile avute – n.a.) sau că este vorba despre rezultatele unui studiu privind eficienţa unei tehnici de meditaţie asupra dezvoltării personale, prin apelul la holism avem acelaşi tip de discurs. Dacă datele observate confirmă presupoziţiile avute, înseamnă că tehnica proiectivă sau meditaţia respectivă sunt tehnici valoroase. Dacă datele observate infirmă aşteptările avute, înseamnă doar că „informaţiile oferite de testele proiective trebuie interpretate contextual şi dinamic, niciodată izolat de istoricul pacientului şi de dinamica elementelor din structura proiectivă analizată”, respectiv că „beneficiile meditaţiei au fost deja demonstrate ştiinţific, doar că, în acest caz, o serie de limite prezente în studiul realizat au afectat posibilitatea de a surprinde eficienţa acestei metode terapeutice”.

Holismul primeşte chiar o coloratură spirituală în multe dintre pseudoştiinţele aparţinând medicinei alternative. În multe dintre aceste ştiinţe există trei idei de bază: (a) funcţionarea normală a energiei vitale asigură o stare de sănătate; (b) blocajul fluxului de energie conduce la o

serie de afecţiuni; (c) deblocarea fluxului de energie vitală va restabili starea de sănătate (Bayerstein, 1996). De pildă, în acupunctură, acele sunt folosite pentru a debloca anumite canale energetice (meridiane). Dacă utilizarea tehnicii are efect este o dovadă a faptului că s-ar fi reuşit deblocarea canalelor energetice, astfel încât fluxul vital să funcţioneze. În schimb, dacă utilizarea tehnicii nu are eficienţa scontată, semnifică faptul că problema este mult mai serioasă, că blocajul energetic este unul puternic, necesitând un timp şi un efort îndelungat pentru deblocare. Se observă, aşadar, că în spatele acestui îndemn la o abordare holistă se regăseşte aceeaşi problemă a lipsei falsificabilităţii, aceasta fiind principala caracteristică distinctivă între ştiinţă şi pseudoştiinţă.

1.3.8. Sofismul ad antiquitatem

Acest sofism porneşte de la ideea greşită că, dacă o idee a reuşit să supravieţuiască un interval lung de timp, fără a fi abandonată, însemnă că este o idee valoroasă. „Vechii învăţaţi chinezi”, „înţelepciunea pierdută a vechilor civilizaţii” sunt numai câteva dintre formulările care încearcă să adauge forţă unor argumente prezentate, prin rezistenţa acestora în timp. Abordarea New Age a încercat să găsească similarităţi între scrierile de filosofie orientală antice şi descoperirile recente din fizica cuantică. Însă, aşa cum remarca Bayerstein (1996) interpretarea liberă a acestor similitudini se bazează pe acelaşi gen de speculaţii precum cele întâlnite în căutarea de similitudini dintre scrierile lui Nostradamus şi evenimentele petrecute de-a lungul timpului, pornind de la înclinaţia naturală a individului de a căuta semnificaţii pentru un anumit eveniment petrecut mai mult decât ar fi cazul (Hyman, 1977).

Caseta 1.4. *Ilustrarea cumulativă a unor elemente descriptive specifice pseudoştiinţelor (adaptare după Marin, 2009)*

Eneagrama este un instrument psihologic foarte performant cu originea în înţelepciunea antică ce descrie nouă stiluri sau tipuri de personalitate, de fapt nouă stiluri diferite de a ne trăi viaţa. Ce are unic eneagrama este faptul ca este singurul sistem psihospiritual care descrie personalitatea umană în totalitatea ei, ca un continuum de stări şi niveluri de conştiinţă, dezvăluind misterul stărilor contradictorii pe care le trăim. Este o sinteză a celor mai noi teorii psihologice: teoria personalităţii multidimensionale, teoria relaţiei cu obiectul, teoria actualizării sinelui, dar şi a vechilor tradiţii esoterice. Chiar simbolul eneagramei se crede că este un simbol de origine pitagoreică datorită proprietăţilor sale matematice speciale.

1.3.9. Aserţiuni generale şi/sau exagerate

Tehnicile propuse de o bună parte dintre curentele ezoterice sau dintre abordările pseudoştiinţifice nu sunt unele oarecare, ci unele care neapărat aduc „o schimbare profundă în viaţa ta, o provocare enormă, vor revoluţiona sistemul tău de valori şi te vor ajuta să fii cu adevărat liber”. De exemplu, terapia câmpului gândirii propusă de Callahan (2001) este o formă de psihoterapie considerată a fi pseudoştiinţifică, conform căreia gândurile şi experienţele trăite se transformă într-un câmp al gândirii, în cadrul căruia orice deformare semnifică o problemă emoţională. Pentru a înlătura problemele noastre emoţionale e nevoie să îndreptăm deformarea câmpului gândirii, iar acest lucru se poate face prin mijloace tactile, dacă se apasă asupra anumitor meridiane energetice pentru a restabili echilibrul lui *chi* (*qi* – energia vitală în cultura tradiţională chineză). Pe lângă problematica în sine, legată de teoria ezoterică aflată la baza acestei forme de psihoterapie, există şi pretenţia susţinătorilor ei că intervenţia este utilă pentru orice fel de tulburare psihopatologică, ea putând fi aplicată nu numai la adulţi, ci şi la copii sau animale.

Însă nu toate abordările pseudoştiinţifice se rezumă la a oferi aserţiuni extravagante. Dimpotrivă, unele abundă de aserţiuni general valabile sau de aserţiuni aparent valide. În această categorie intră abordările care apelează frecvent la un discurs bazat pe efectul Barnum (aserţiuni general valabile, dar lipsite de valoare predictivă la nivel individual) (Forer, 1949 *apud* Pelham şi Blanton, 2007). De pildă, dacă cineva vă prezintă rezultatele unui test psihologic, conform cărora aflaţi că „sunteţi o persoană care se simte bine atunci când sunteţi apreciată de către ceilalţi, că uneori aveţi tendinţa de a fi mai critic(ă) cu dvs., că vă frustrează încercările celorlalţi de a vă impune anumite limite sau restricţii, că preferaţi un anumit grad de schimbare în viaţa dvs., dar fără a renunţa la valorile şi principiile care vă definesc ca persoană", veţi afirma, foarte probabil, că instrumentul utilizat surprinde corect descrierea modului dvs. de a fi. Problema acestui gen de afirmaţii este că este general valabil, prin urmare nefalsificabil, deci neinclus într-o abordare ştiinţifică în psihologie. Acest tip de afirmaţii general valabile conduce adeseori la proliferarea unui discurs ştiinţific de tipul „bunului-simţ", la care recurg mulţi specialişti în încercarea lor de a oferi interpretări sau explicaţii pentru evenimente/fenomene psihologice pe care nu le cunosc foarte bine. Astfel, revolta tinerilor este mereu un rezultat al nevoii lor nesatisfăcute de a fi trataţi de la egal la egal, ca nişte adulţi, iar problemele emoţionale ale copiilor având părinţii plecaţi la muncă în străinătate au drept sursă deficitul de afecţiune cu care ei se confruntă etc.

O categorie specială de interpretări specifice simţului comun este apelul la semnificaţia simbolică. În Antichitate, Aristotel argumenta că gravitaţia reflectă dorinţa obiectelor de a se întoarce către pământul-sursă, la locul de care aparţin (Rensberger, 1986 *apud* Pelham şi Blanton, 2007). Desigur, fizica a avansat mult şi a renunţat la acest gen de explicaţii de

sorginte animistă. Nu acelaşi lucru se poate spune despre psihologie, care se confruntă încă cu acest gen de interpretări simbolice, chiar dacă nu de sorginte animistă. De pildă, în grafologie, un scris mic şi înghesuit poate fi un semn al reticenţei sau zgârceniei persoanei respective, în timp ce un scris neîngrijit, cu multe erori, reflectă stilul expansiv şi impulsiv de a acţiona. În alte domenii, cum ar fi comunicarea nonverbală, poziţia unui individ cu braţele încrucişate indică opoziţia şi reticenţa acestuia faţă de mesajul transmis, în timp ce unele teorii psihanalitice consideră actele de gelozie nimic altceva decât o expresie a neîncrederii în capacitatea propriei persoane de a rămâne fidel(ă) faţă de partenerul său. La prima vedere, interpretările pot părea plauzibile, însă ele suferă puternic în ceea ce priveşte validitatea lor empirică, predictivă.

1.3.10. Abuzul de limbaj tehnic (specializat)

În încercarea de a se apropia cât mai mult de abordările ştiinţifice, unele tehnici pseudoştiinţifice au împrumutat limbajul şi discursul ştiinţific pentru a-şi expune teoria. De pildă, Herbert *et al.* (2000; vezi caseta 1.5 mai jos) prezintă modul în care autorii unei alte forme de psihoterapie considerată ca fiind pseudoştiinţifică – EMDR (*Eye Movement Desensitization and Reprocessing*) îşi prezintă modelul procesării accelerate a informaţiei ca baza teoretică a intervenţiei într-o manieră ce pare credibilă.

Aşadar, este doar un stereotip faptul că pseudoştiinţele ar respecta patternul holismului şi al căderii în admiraţie pentru filosofia orientală antică. Ele pot disimula foarte bine, utilizând un discurs ştiinţific elevat. De notorietate rămâne farsa jucată de fizicianul Sokal care a reuşit să publice un text fără sens în revista postmodernistă *Social text*, primind

acceptul evaluatorilor de a publica articolul cu titlul original „Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity”, din cadrul căruia este prezentat un mic pasaj ca notă de subsol, păstrând varianta originală a textului, în limba engleză[2], reluat de Lilienfeld şi Landfield (2008, p. 1225).

Caseta 1.5. *Exemplu de abuz în utilizarea unui limbaj tehnic (Herbert* et al.*, 2000, p. 960)*

„Receptorii neuronali (potenţialul sinaptic) incluşi în reţelele neuronale respective conţin niveluri diferite ale pragurilor de activare, denumite de la A la Z, în funcţie de valoarea adaptivă a informaţiei encodate. Se presupune că reţelele neuronale ţintă Z au valori prea ridicate ale pragurilor de activare, ceea ce conduce la imposibilitatea conectării acestor reţele la alte reţele ce conţin informaţii cu valoare adaptivă. Cele din urmă pot fi activate mai uşor, datorită valorii scăzute a pragurilor de activare. Astfel, potenţialul sinaptic este diferit, pentru fiecare tip de afect encodat în diferite reţele neuronale (pragurile de activare fiind cele mai ridicate pentru acele reţele neuronale care encodează trăirile afective asociate evenimentului traumatic – n.a.)*… Prin intermediul EMDR, pragul de activare a receptorilor neuronali din reţele scade, putându-se stabili conexiuni cu alte reţele ce conţin informaţie adaptivă, ale căror prag de activare neuronală este mai scăzut*”.

1.3.11. Ideea evoluţiei/superiorităţii adepţilor

Abordarea ştiinţifică nu conţine în discursul său referiri de ordin calitativ cu referire la adepţii uneia sau alteia dintre teoriile ştiinţifice. Unele pseudoştiinţe tind să facă o asemenea diferenţă între cei care îmbrăţişează teoriile lor şi cei care nu cred în aceste idei. Primii sunt cei care se consideră „evoluaţi spiritual”, „deschişi spre adevăr şi spre a surprinde esenţa Universului”, „pregătiţi pentru progres” „integraţi spiritual pe calea cunoaşterii sinelui profund, dincolo de orice iluzie a lumii materiale”. Ceilalţi se află încă la un nivel „inferior”, „nepregătiţi să facă pasul spre progresul spiritual”, „prea rigizi şi dogmatici, reducţionişti”, „incapabili

să sesizeze complexitatea universului şi legăturile ascunse, esenţele ultime".

1.3.12. Absenţa conexiunilor cu teoriile anterioare şi cu celelalte ştiinţe

În abordarea ştiinţifică, discursul se construieşte pas cu pas, pornind de la teoriile deja existente, aducându-li-se acestora corecţii, completări sau noi perspective pe baza unei analize critice. Aceasta deoarece pentru a fi considerată ca un progres în cunoaştere, o teorie ar trebui să explice nu numai aspectele noi, ci să analizeze (integreze) modelele mai vechi din prisma noii teorii. În multe dintre abordările pseudoştiinţifice lipseşte această conexiune cu teoriile anterioare. Apare în schimb ideea că abordarea este una revoluţionară, ce necesită noi paradigme asupra realităţii. Chiar dacă asemenea situaţii au apărut rar şi în domeniul ştiinţific (vezi trecerea la teoria relativităţii a lui Einstein), ele apar mult mai frecvent ca aserţiuni în justificarea unor construcţii teoretice din domeniul pseudoştiinţelor.

În plus, un atu important al unei teorii ştiinţifice este nu numai conexiunea cu teorii anterioare din domeniu, ci şi conexiuni dovedite ştiinţific cu alte ştiinţe. În acest sens teoria szondiană a genelor pulsiunilor şi a destinului este vetustă dacă ne raportăm la rezultatele obţinute din cercetările genetice, lipsind fundamentarea biologică pentru o asemenea teorie.

Aşadar, în pofida dezvoltării extraordinare a interesului pentru psihologie, este nevoie de multă atenţie din partea comunităţii ştiinţifice psihologice de a nu cădea pradă abordărilor de tip pseudoştiinţific. Nevoia societăţii contemporane de a găsi explicaţii psihologizante pentru o serie

de evenimente şi situaţii cu care se confruntă avantajează, paradoxal, abordările de tip pseudoştiinţific, îndeosebi ca urmare a câtorva dintre caracteristicile acestora: (a) oferirea de explicaţii post eveniment prin intermediul unor teorii nefalsificabile; (b) mantaua holismului şi a legăturii cu spiritualitatea; (c) accentul pus pe confirmare prin apelul la opinii şi cazuri particulare, ce sunt mult mai uşor de obţinut decât dovezile experimentale.

1.4. Consecinţe negative ale abordării pseudoştiinţifice în psihologie

Riscurile derivate din popularitatea curentelor pseudoştiinţifice printre consumatorii de psihologie sunt cu atât mai mari, cu cât în alte domenii apare o presiune inversă, de fundamentare ştiinţifică a oricărei practici profesionale, cel mai ilustrativ exemplu fiind abordarea validării ştiinţifice venită dinspre medicină (*evidence based medicine* – EBM).

Astfel, un prim risc major ce apare în absenţa unei abordări ştiinţifice în psihologie este unul de ordin financiar. De pildă, serviciile psihologice din zona asistenţei medicale se află în competiţie pentru fonduri cu alte domenii din sfera asistenţei medicale, inclusiv farmacoterapia. Din acest punct de vedere, nevoia de servicii psihologice validate ştiinţific este una acută de vreme ce, deşi cheltuielile, în termeni procentuali, pentru servicii de sănătate mentală au rămas aproximativ constante în ultimii 30 de ani, dinamica costurilor în interiorul acestor servicii a suferit modificări semnificative. Astfel, costurile alocate pentru farmacoterapie au crescut, în timp ce cheltuielile pentru servicii de psihoterapie au scăzut, dacă ne raportăm în termeni procentuali (APA Task Force, 2009), iar acest aspect

constituie un semnal de alarmă mai cu seamă că numărul de psihologi a crescut de patru ori în ultimii 50 de ani (Goodheart, 2010). Aspectul principal care a condus la această stare de fapt a fost acela că, în lipsa unor resurse financiare suficiente, casele de asigurări au început să plătească doar acele servicii pentru care există dovezi de validitate ştiinţifică. Cum domeniul farmacoterapiei a îmbrăţişat înaintea psihoterapiei ideea unei abordări bazată pe validarea ştiinţifică, constatăm o creştere a acestui tip de cheltuieli în dauna psihoterapiei, în cadrul categoriei de cheltuieli privind sănătatea mentală.

În strânsă legătură cu acest aspect, apare ideea creşterii responsabilităţii profesionale (*accountability*) în domeniul evaluărilor şi intervenţiilor psihologice. În alte domenii din sfera sănătăţii mentale, o consecinţă importantă a abordării bazate pe validarea ştiinţifică a fost aceea a standardizării intervenţiilor prin dezvoltarea unor protocoale (proceduri standard de lucru). Acest aspect a permis tragerea la răspundere a celor care nu respectau metodologia de lucru, un aspect important atât pentru companiile de asigurări, cât şi pentru casele de sănătate, în demersul lor de a deconta serviciile prestate. Numai o abordare bazată pe validare ştiinţifică va putea dezvolta, în timp, protocoale standard de intervenţie în cadrul psihologiei şi al psihoterapiei, care să permită evaluarea respectării responsabilităţii profesionale a specialiştilor din psihologie şi psihoterapie.

Un al treilea risc pe care îl presupune o abordare de tip pseudoştiinţific în sfera psihologiei este acela al izolării şi marginalizării domeniului, cel puţin în sfera serviciilor de sănătate mentală. În contextul în care se tinde către oferirea de servicii integrate de sănătate mentală (APA Task Force, 2009; Spring, 2007) şi în care celelalte servicii din sfera medicală sunt fundamentate ştiinţific (psihiatria, farmacoterapia, asistenţa

medicală – *nursing*), este necesar ca şi intervenţiile de tip psihologic să fie validate empiric, pentru a putea fi incluse în acest pachet integrat de servicii de sănătate mentală.

Acestor riscuri li se adaugă cel al percepţiei distorsionate a domeniului psihologiei, în contextul în care mulţi dintre absolvenţii de psihologie nu ţin cont în alegerea formării lor profesionale continue de perspectiva bazată pe validare ştiinţifică. În următorul capitol al acestei părţi introductive sunt prezentate în detaliu caracteristicile unei practici validate ştiinţific în domeniul psihologiei. Vom observa, cât de puţine dintre elementele prezentate acolo sunt incluse în practica uzuală a specialiştilor din psihologie şi psihoterapie. Din nefericire, discursul majorităţii psihologilor, cel puţin din România, constituie încă un amalgam format din clişee lingvistice bazate pe simţul comun („copiii trebuie ajutaţi să-şi dezvolte încrederea în forţele proprii") şi din opinii nevalidate empiric, preluate într-o manieră scolastică, pe parcursul formării profesionale continue („deoarece, aşa cum afirma Freud, pulsiunile sexuale constituie motorul comportamentului uman").

1.5. Descriptori ai unei abordări ştiinţifice

Un element extrem de important de reţinut este acela că nu conţinutul studiat este cel care determină diferenţa dintre ştiinţă şi pseudoştiinţă. De pildă, nu faptul că cineva decide să studieze fenomenele extrasenzoriale (premoniţii, precogniţii), adică subiecte contradictorii pentru lumea ştiinţifică, determină caracterul ştiinţific sau pseudoştiinţific al acestui demers. Ceea ce constituie diferenţa este apelul la metoda ştiinţifică. Dacă

studiile efectuate utilizează o metodă ştiinţifică şi trec cu succes prin mecanismul de *peer-review* în cadrul unor reviste relevante pentru un anumit domeniu studiat, putem considera un asemenea demers ca fiind unul ştiinţific (vezi, de exemplu, studiul lui Bem (2011) despre precogniţii şi premoniţii publicat în prestigioasa revistă *Journal of Personality and Social Psychology*)[3]. În schimb, dacă studiile sunt caracterizate mai degrabă de indicatori specifici abordării pseudoştiinţifice, inclusiv acela de evitare a procesul de *peer-review* sau de simulare a acestuia în cadrul unor reviste marginale, de „cas(t)ă”, putem vorbi despre un demers pseudoştiinţific.

În bună măsură, indicatorii unui demers ştiinţific, pot fi identificaţi ca opusul celor specifici pseudoştiinţelor. Astfel, abordarea ştiinţifică se bazează ca mecanism de funcţionare pe principiul falsificabilităţii, pe evitarea biasului de confirmare, pe existenţa revizuirilor teoretice etc. Stanovich (2010) surprinde patru caracteristici definitorii pentru orice demers ştiinţific prezentate, pe larg, în cele ce urmează. La acestea aş adăuga o altă caracteristică esenţială, aceea de a adopta o atitudine sceptică faţă de rezultatele obţinute în orice studiu, oferind o însemnătate deosebită explicaţiilor alternative.

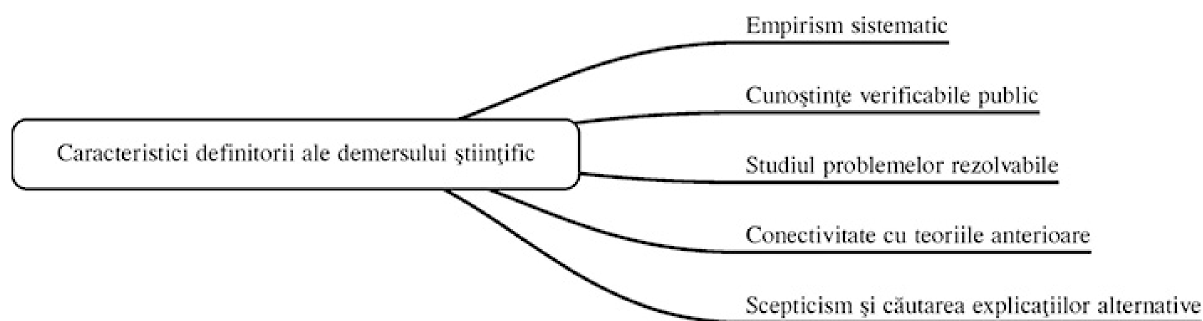


Figura 1.2. *Caracteristici definitorii ale unui demers ştiinţific*

1.5.1. Empirismul sistematic

Oamenii manifestă psihologic înclinaţia de a face predicţii sau de a oferi explicaţii şi interpretări pentru evenimentele din viaţa lor. Însă, în acest demers ei apelează rar la un demers ştiinţific, bazat pe ceea ce se numeşte empirism sistematic (Gravetter şi Forzano, 2009; Stanovich, 2010). În schimb, ei îşi bazează explicaţiile pe *credinţe populare* (de exemplu, „să nu stai pe colţul mesei pentru că nu te vei căsători") sau pe *intuiţii personale* rezultate în urma experienţei personale care le creează senzaţia că au dreptate (de exemplu, o femeie gravidă care afirmă: „sunt convinsă că voi naşte o fetiţă pentru că simt sarcina diferit faţă de prima dată când am născut un băiat"). Este evident faptul că asemenea elemente nu pot face parte dintr-o abordare de tip ştiinţific.

O altă sursă care stă la baza modului lor de cunoaştere şi interpretare a lumii se referă la *apelul la autoritate*, fie aceasta una de ordin ştiinţific, filosofic sau religios. În această categorie intră informaţiile furnizate de experţii dintr-un domeniu. Cu toate că metoda este deosebit de economică în ceea ce priveşte consumul de timp sau de alte resurse, există un risc ridicat ca unele dintre aceste opinii să fie eronate. De aceea, apelul la autoritate nu este considerat un element acceptabil din perspectiva unui demers ştiinţific, indiferent de prestigiul şi credibilitatea celui care a emis o anumită afirmaţie. Acest lucru se întâmplă pentru că aceste opinii pot fi distorsionate (sau chiar eronate din punct de vedere ştiinţific, cum ar fi cazul explicaţiei deficitului de afectivitate matern ca o cauză pentru apariţia autismului emisă de psihanalistul Bruno Betellheim [Wikipedia, 2009]). Mai mult, presiunile exercitate de grupurile de *advocacy* cresc riscul emiterii unor opinii distorsionate care să susţină o cauză sau alta, în funcţie de interesul grupului de suport (de exemplu, vezi opiniile pro sau împotriva adopţiei internaţionale sau cele ale adopţiilor de către familii/cupluri de acelaşi sex). Nu în ultimul rând, un alt neajuns al

fundamentării cunoaşterii pe deciziile unor experţi este acela al confuziei domeniului de expertiză. Să ne înţelegem, doar pentru că o persoană a avut contribuţii importante într-un domeniu, nu înseamnă că mereu are dreptate. Wundt, întemeietorul psihologiei ştiinţifice prin intermediul laboratorului de la Leipzig, nominalizat de două ori pentru acordarea premiului Nobel pentru medicină (Academia Premiilor Nobel, 2012), se declara împotriva votului universal („Franţa… şi-a săpat mormântul prin reforma votului universal", Wundt, 1866 *apud* Bringmann, Ungerer şi Bringmann, 1995, p. 239). Tot el scria materiale de propagandă antiamericane şi antibritanice în Primul Război Mondial (New World Encyclopedia, 2008). În mod similar, numeroşi intelectuali români apreciaţi, inclusiv în zilele noastre, au făcut parte din ceea ce Eugen Ionescu denumea plastic fenomenul de rinocerizare a intelectualilor (*apud* Morar, 2005). Mircea Eliade a îmbrăţişat şi susţinut legionarismul, Mihai Eminescu, Nicolae Iorga, Constantin Noica sau Emil Cioran au scris numeroase texte cu conţinut antisemit, iar ultimul a mers chiar mai departe, aducând elogii lui Hitler şi lui Mussolini, precum şi ideii de stat totalitar (de exemplu, „Nu există om politic în lumea de astăzi, care să-mi inspire o simpatie şi o admiraţie mai mare decât Hitler", în *Vremea* nr. 346, din 15 iulie 1934 *apud* Morar, 2005). De aceea este deosebit de important să reţinem că, din punct de vedere ştiinţific, valoarea unei idei sau opinii nu este judecată după CV-ul celui care o enunţă şi nici nu se transferă automat cu valoarea intelectuală a enunţătorului, deşi există înclinaţia retorică, întâlnită în cazul unora, de a-şi afişa toate titlurile şi distincţiile obţinute, înainte de a exprima o opinie, pentru a da un plus de greutate acesteia. Mai mult, opiniile nu devin valide ştiinţific doar bazându-se pe prestigiul enunţătorului, chiar dacă ele vizează probleme din cadrul domeniului de expertiză al acestuia. În acest sens, destul de

cunoscută este afirmaţia eronată făcută de către Einstein, în 1932, când afirma că sunt puţine indicii care să arate că se va putea obţine vreodată energie nucleară.

O altă sursă majoră de cunoaştere este *apelul la raţionamentele logice*. Desigur, dacă premisele sunt adevărate şi dacă raţionamentul este unul corect din punct de vedere logic, cu siguranţă concluzia este una adevărată. Din păcate, în realitate asistăm de multe ori la formularea de raţionamente pornind de la una sau mai multe premise eronate sau eşuăm în a identifica greşeli în modul de constituire a silogismelor. Una dintre cele mai întâlnite erori privind raţionamentele în domeniul ştiinţific este sofismul afirmării consecinţei.

Dacă X, atunci Y
Y
Deci, X

Transpus în termeni ştiinţifici acest gen de raţionament devine:

Dacă teoria X este adevărată, atunci vom anticipa că obţinem Y;
obţinem Y;
deci, teoria X este adevărată.

Acest raţionament este unul greşit deoarece Y poate fi explicat, probabil, prin alte teorii alternative. În exemplul: (a) „dacă a plouat recent, pământul este umed”; (b) „pământul este umed”; (c) „deci, a plouat recent” putem găsi şi alte explicaţii pentru care pământul ar fi umed (cum ar fi aceea de a fi fost stropit de către un grădinar). În schimb, raţionamentul ar fi unul corect dacă ar lua o altă formă logică, cum ar fi

negarea consecinţei, precum în exemplul: (a) „dacă plouă recent pământul este umed”; (b) „pământul nu este umed”; (c) „deci, nu a plouat recent”.

Dacă X, atunci Y
Non Y
Deci, non X

Transpus în logica ştiinţifică, raţionamentul corect este:

Dacă teoria X este adevărată, atunci vom anticipa că vom obţine Y;
nu obţinem Y;
deci, teoria X nu este adevărată.

Pentru a complica şi mai mult lucrurile, chiar asemenea argumente construite corect pot ajunge în practică la concluzii greşite deoarece se pot strecura alte erori, de ordin metodologic, care să inhibe apariţia efectului Y. În practică putem observa că pământul nu este umed, chiar dacă a plouat recent, dacă pământul avut sub observaţie este aşezat sub o copertină. În mod similar, în studii de specialitate se pot strecura erori de ordin metodologic, care conduc la rezultate distorsionate. În aceste condiţii nu înseamnă neapărat că teoria din spatele studiului este una greşită, ci este posibil ca anumite artefacte metodologice să conducă la eşecul observării efectului.

În plus, capacitatea de a raţiona corect, esenţială pentru a ajunge la concluzii adevărate, este afectată în cazul multora dintre oameni, care se lasă înşelaţi de aparenţe. De exemplu, 67% dintre studenţii de masterat chestionaţi recent au afirmat că raţionamentul prezentat în partea stângă a casetei 1.6 ar fi unul corect, în timp ce niciunul dintre masteranzi nu a

considerat raţionamentul din partea dreaptă a aceleaşi casete ca fiind unul corect, deşi ambele au aceeaşi structură conform logicii propoziţionale. Soluţia formală la acest gen de raţionamente poate fi oferită doar prin apelul la logica predicatelor.

Toţi X sunt M
Unii M sunt Y
Unii X sunt Y

Caseta 1.6. *Exemplu de raţionament, identic propoziţional, care duce la concluzii diferite*

Toate femeile sunt fiinţe umane	*Toate femeile sunt fiinţe umane*
Unele fiinţe umane sunt psihologi	*Unele fiinţe umane sunt bărbaţi*
Deci, unele femei sunt psihologi	*Deci, unele femei sunt bărbaţi*

O ultimă sursă importantă de cunoaştere este *metoda empirică*, bazată pe ideea de a colecta date despre situaţia existentă prin simpla observaţie. Astfel, putem cunoaşte, de exemplu, că fiinţa umană are patru membre, două inferioare şi două superioare; că iarna nu este la fel de cald ca vara; că există un ciclu periodic al anotimpurilor etc. Dar apelul exclusiv la observaţii şi date empirice, poate fi iluzoriu, aproape 50% dintre români cred că Soarele se învârte în jurul Pământului şi nu invers (Vlăsceanu, Duşa şi Rughiniş, 2010), iar un număr din ce în mai însemnat de copii cred că vacile au culoarea mov, preluând reprezentarea din celebra marcă de ciocolată (FDSC, 2010). Nu în ultimul rând, aceste observaţii empirice pot fi obiectul unor interpretări eronate, bazate pe credinţele, aşteptările şi cunoştinţele anterioare ale celui care face observaţia, ceea ce constituie o altă limită a cunoaşterii bazate exclusiv pe metoda empirică.

Ca urmare, toate sursele de informaţii prezentate anterior, fie că este vorba de intuiţii sau credinţe populare, de opinii ale unor autorităţi

ştiinţifice sau religioase, de demersuri raţionale, logice sau de colectarea de date empirice, au o serie de limite care pot conduce la distorsionarea cunoaşterii. Niciuna dintre ele nu se suprapune cu empirismul sistematic întâlnit într-o abordare ştiinţifică a cunoaşterii. Însă ambii termeni conţinuţi în acest construct de empirism sistematic sunt definitorii. Astfel, empirismul indică predispoziţia unui demers ştiinţific de a colecta date din realitate (empirism), cu precizarea că acest demers este unul sistematic (controlat, structurat), orientat spre testarea unei teorii. Prin controlul efectuat asupra metodelor empirice de colectare a datelor se doreşte evitarea erorilor care ar conduce la obţinerea de date distorsionate despre realitate, iar prin caracterul structurat şi logic al acestui demers se are în vedere realizarea măsurătorilor (observaţiilor) într-un context specific, al testării unor teorii sau a unor modele, care să permită ulterior realizarea de inferenţe (predicţii, explicaţii) despre modul de funcţionare a realităţii, dincolo de ceea ce este imediat şi direct observabil, măsurabil. Prin urmare, empirismul sistematic nu se rezumă la simpla observare a realităţii, la a colecta date, ci are în vedere un demers sistematic de verificare a unei teorii (ipoteze, predicţii)[4]. De aceea, cea mai mare parte dintre studiile realizate conţin cel puţin o ipoteză ce trebuie cercetată şi o teorie pe care aceasta se bazează (testarea unei teorii nefiind altceva decât demersul de generare şi de testare a unui set de ipoteze corelate, ce derivă din teoria respectivă).

De aceea, prin empirism sistematic (vezi, spre ilustrare, caseta 1.7), vom înţelege un demers complex ce cuprinde: (a) observarea realităţii prin prisma unei teorii (lentile de descifrare a realităţii); (b) generarea unei ipoteze cu scopul de a verifica realitatea observată; (c) testarea ipotezei printr-un demers investigativ adecvat; (d) finalizarea demersului

printr-un răspuns care poate sprijini sau nu teoria formulată, într-o manieră indirectă, prin intermediul ipotezei generate din teoria respectivă.

Un lucru foarte important legat de acest demers sistematic empiric este faptul că niciodată în ştiinţă nu este corect să utilizăm expresii de genul „ipoteza este confirmată”, „teoria este dovedită (a fost demonstrată)”, ci doar expresii de tipul „ipoteza este susţinută”, „teoria este sprijinită de datele empirice colectate”. Acest lucru se întâmplă pentru a evita sofismul afirmării consecinţei prezentat anterior. Astfel, un rezultat ştiinţific conform căruia frecvenţa mai ridicată a viselor cu conţinut violent întâlnită în cazul băieţilor, comparativ cu cel al fetelor, poate fi interpretat diferit, atât ca un sprijin în favoarea teoriei psihanalitice a complexului lui Oedip (teama băieţilor de a fi castraţi de tatăl lor ca urmare a dragostei lor pentru mamă), cât şi ca o consecinţă a unor diferenţe de gen existente la nivel genetic (Coolican, 2004).

Pe scurt, prin acest demers empiric sistematic, ştiinţa avansează prin formularea de teorii explicative pentru anumite fenomene observate, pe baza cărora se pot face o serie de predicţii a căror valoare de adevăr este testată prin demersul ştiinţific realizat. În funcţie de rezultatele obţinute se poate ajunge, fie la sprijin pentru teoria respectivă şi la identificarea unor legităţi cu valabilitate largă, fie, dimpotrivă, la revizuirea teoriei respective, pentru a putea include rezultatele obţinute sau chiar la eliminarea teoriei respective, în cazul în care noile date obţinute nu pot fi explicate nici măcar de o variantă revizuită a teoriei respective.

Caseta 1.7. *Paşii unui demers empiric sistematic. Ilustrarea unor cazuri ipotetice de studiu*

Paşii de lucru	Exemplul 1	Exemplul 2
Pasul 1.	În ultimele luni de zile am	Se constată adesea, că mulţi dintre indivizii care

Observarea realităţii prin prisma unei teorii	crescut în greutate. Am observat, în aceeaşi perioadă, că am renunţat să mai practic orice activitate fizică intensă (asociere pe bază de inducţie). Este foarte probabil ca surplusul de calorii neconsumate să fie cauza creşterii în greutate (deducţie pe bază teoretică – cunoştinţe anterioare din domeniul nutriţiei).	renunţă la locul lor de muncă actual fac parte dintre cei care percep munca depusă ca fiind una extrem de solicitantă (lucrează mult, greu, în condiţii de stres), fiindu-le greu să facă faţă cerinţelor postului. Din perspectivă teoretică, posturile solicitante sunt cele în care se intensifică riscul de a dezvolta probleme de sănătate fizică şi mentală printre angajaţi, mai cu seamă de către angajaţii care dispun de resurse puţine.
Generarea ipotezei de lucru[1]	Demararea unui program de activităţi fizice va reduce/elimina tendinţa de a creşte în greutate (comparaţie între situaţia greutăţii mele în lunile în care nu am făcut sport *vs* lunile în care am făcut mişcare fizică intensă).	Intenţia unui individ de a părăsi actualul loc de muncă se află în legătură cu nivelul de solicitare resimţit faţă de postul pe care îl ocupă. Un nivel ridicat de solicitare resimţit pe post este asociat cu un risc mai ridicat al fluctuaţiei de personal.
Testarea ipotezei de lucru	În condiţiile menţinerii aceluiaşi regim alimentar, pentru a controla inputul energetic primit de organism, se colectează lunar date cu privire la greutate şi se compară diferenţele în greutate sesizate între lunile cu şi cele fără activitate sportivă.	Un lot de muncitori (indivizi cu autonomie decizională scăzută) este investigat prin intermediul a două chestionare standardizate şi validate din punct de vedere ştiinţific, ce măsoară gradul de solicitare resimţit pe post, respectiv nivelul intenţiei de a părăsi organizaţia în care lucrează.
Integrarea răspunsului din perspectiva unei teorii	S-a constatat o creştere în greutatea corporală în lunile fără activitate fizică intensă, fapt ce sprijină teoria surplusului de calorii neconsumate ce stă la baza creşterii în greutate. Astfel,	S-a constatat absenţa unei asocieri între intenţia de a părăsi compania şi nivelul de solicitare pe post. În lipsa unor explicaţii ce vizează limite metodologice, un asemenea rezultat ar trebui să conducă la revizuirea modelului teoretic conform căruia suprasolicitările postului (*job strain*) constituie un factor de risc în apariţia fluctuaţiei de personal. De exemplu, în studii

	activitatea fizică este una din modalităţile de echilibrare calorică.	viitoare ar putea fi căutaţi factori moderatori de care depinde apariţia efectului menţionat în teorie, cu alte cuvinte să se arate că efectul suprasolicitării asupra fluctuaţiei de personal apare doar în anumite condiţii (de exemplu, un venit lunar nesatisfăcător).

1.5.2. Cunoştinţe verificabile public

După empirismul sistematic, cea de a doua caracteristică majoră a oricărei abordări ştiinţifice este aceea de a permite verificarea publică a demersului realizat şi a rezultatelor obţinute.

Procedura de bază utilizată în acest sens este replicarea studiilor anterioare (Stanovich, 2010). Pentru ca acest lucru să se întâmple, orice demers ştiinţific este descris suficient de detaliat încât să permită altor persoane interesate să replice studiul pentru a urmări măsura în care sunt obţinute rezultate similare. Această măsură poate fi privită ca o modalitate de protecţie împotriva erorilor apărute în procesul de testare a ipotezelor, fie ele erori de tip I (rezultate semnificative statistic, când nu există un efect), de tip II (rezultate nesemnificative statistic, când există un efect), de tip III (rezultate semnificative statistic, dar în sens contrar celui aşteptat). De asemenea, prin replicarea rezultatelor se oferă o siguranţă suplimentară că rezultatele nu sunt distorsionate de existenţa vreunor probleme metodologice. În cadrul procesului de replicare ideal ar fi ca demersul de replicare să fie realizat de o altă echipă de cercetare decât cea care a descoperit iniţial efectul studiat, pentru a asigura o mai mare încredere în rezultatul replicat.

O altă modalitate esenţială de verificare publică a cunoştinţelor generate prin cercetări este aceea de a publica rezultatele în reviste de specialitate în care funcţionează evaluarea de tip colegial (*peer-review*). Chiar dacă acest proces nu este lipsit de subiectivitate, acesta constituie

cel mai bun mijloc existent de a evalua critic corectitudinea demersului ştiinţific efectuat şi validitatea rezultatelor obţinute. Deşi prin multitudinea de reviste de specialitate existente într-un domeniu (de exemplu, baza de date PsycInfo însumează peste 2.000 de reviste de specialitate – Stanovich, 2010) se pot strecura şi studii cu rezultate invalide (problematice), absenţa supunerii rezultatelor unei cercetări la procesul de evaluare colegială este un semn clar a unui demers neştiinţific. În cazul evaluării de tip *peer-review* rezultatele dobândesc o forţă ştiinţifică mai mare dacă sunt publicate: (a) în reviste ISI Web of Science de top, cu factor de impact ridicat pentru domeniul respectiv; (b) în reviste ce nu pot fi bănuite ca fiind revistele casei (de exemplu, un articol favorabil unei teorii noi să fie publicat într-o revistă distinctă de cea care este condusă chiar de autorul articolului respectiv sau de iniţiatorii şi/sau susţinătorii respectivei teorii).

În sfârşit, un alt aspect relevant ca urmare a introducerii principiului verificării publice a cunoştinţelor este productivitatea crescută a oricărui domeniu ştiinţific, având loc frecvent (aproape cu fiecare deceniu) o revizuire sau înlocuire a teoriilor sau cunoştinţelor existente. Acest progres rapid conduce treptat la consolidarea practicii psihologiei bazată pe validarea ştiinţifică (*evidence based psychology practice*), ceea ce presupune, în linii mari, implementarea de către practicieni a celor mai bune modalităţi de intervenţie pentru a soluţiona diferite aspecte practice, în urma testării eficienţei în studiile de specialitate.

1.5.3. Studiul problemelor rezolvabile

Acum 100 de ani, o problemă precum „reducerea simptomatologiei depresive prin intermediul realităţii virtuale” ar fi fost o temă care nu ar fi

putut fi abordată ştiinţific. Astăzi, datorită progresului ştiinţific, o asemenea temă poate fi abordată ştiinţific. În schimb, o temă precum „studiul efectelor interacţiunii dintre oameni şi fiinţele extraterestre asupra imaginii de sine a fiinţelor umane” nu poate fi considerată a fi o abordare ştiinţifică în prezent, dar ea ar putea deveni o temă respectabilă de studiu în viitor, dacă omenirea va fi marcată în existenţa sa de asemenea evenimente potenţiale. Cu alte cuvinte, pentru a fi considerată ştiinţifică, o abordare trebuie să apeleze la metode de investigaţie care să permită testarea empirică a problematicii studiate. De aceea se poate pune un semn de egalitate între studiul problemelor rezolvabile şi cel de verificare a unor teorii sau modele explicative ce pot fi testate, adică verificate din perspectiva gradului lor de validare empirică. Astfel demersul ştiinţific diferă fundamental de cel pseudoştiinţific prin faptul că abordările pseudoştiinţifice apelează la teorii şi concepte care nu pot fi falsificabile[5].

Un alt aspect important care distinge abordările ştiinţifice de cele neştiinţifice, în direcţia studierii unor probleme rezolvabile, este distincţia dintre operaţionalism şi esenţialism (Stanovich, 2010). Căutarea esenţei ultime (de exemplu, „care este sensul vieţii?”, „ce înseamnă cu adevărat altruismul?”, „care este esenţa moralităţii?”) nu face parte dintr-o abordare ştiinţifică. Abordarea ştiinţifică în această direcţie ar încălca principiul problemelor rezolvabile (nu dispunem, cel puţin momentan, de metode empirice de investigare a acestor aspecte) şi prin urmare nu s-ar supune principiului falsificabilităţii, orice explicaţie oferită rămânând la nivelul unei speculaţii de tip filosofic. Soluţia pentru a aborda ştiinţific o temă este de a renunţa la esenţialism şi de a operaţionaliza conceptele studiate. Ideea de bază a operaţionalismului este aceea că, oricât de abstract ar fi la prima vedere un construct, acesta ar trebui măsurat

empiric printr-un indicator observabil. La prima impresie acest lucru este uşor de realizat, de exemplu, pentru a măsura nivelul inteligenţei generale a unei persoane cineva poate apela la rezultatul global obţinut de persoana respectivă la scalele de inteligenţă Wechsler pentru adulţi – WAIS-IV (Wechsler, 2008). Însă modul de operaţionalizare a unui construct teoretic abstract poate conduce la indicatori variaţi, rezultaţi în urma unor decizii complexe care pot afecta demersul ştiinţific la nivelul validităţii sale de construct (vezi secţiunea 3.4.3). De pildă, pentru evaluarea satisfacţiei maritale se poate apela la un indicator de autoraportare global pe o scală Likert în cinci trepte („Per ansamblu cât de satisfăcut(ă) sunteţi de relaţia pe care o aveţi cu soţia (soţul) dvs.?") sau la un indicator multidimensional bazat pe evaluarea relaţiei dintre parteneri cu privire la mai multe aspecte: sexual, afectiv, de într-ajutorare cu privire la aspectele administrative etc. O altă modalitate de operaţionalizare a gradului de satisfacţie în relaţia de cuplu ar putea fi prin intermediul unei grile de observare completată de o persoană terţă, care are acces să urmărească relaţia de cuplu periodic. Desigur, alte elemente luate în considerare pot afecta răspunsurile celor chestionaţi precum: perioada de raportare (în general, în ultimele şase luni etc.), utilizarea unor ancore facilitatoare cu rol comparativ (de exemplu, una este instrucţiunea de exprimare a gradului de satisfacţie în comparaţie cu celelalte cupluri pe care le cunoaşteţi, alta este exprimarea evaluării în comparaţie cu aşteptările pe care le aveţi de la partener). La acestea se adaugă eventualele limite ale metodelor utilizate, cum ar fi tendinţa celor chestionaţi să ofere răspunsuri dezirabile social. De aceea este important ca în orice demers de operaţionalizare a conceptelor studiate să se bazeze pe proceduri validate anterior (vezi secţiunea 3.4.3 destinată validităţii de construct). Prin definirea operaţională a constructelor uzitate, abordarea ştiinţifică evită

definirea ambiguă, poetică a unor concepte atât de sensibile în psihologie precum inteligenţa, dragostea, performanţa etc. Totuşi, acest demers pragmatic nu este lipsit de critici, precum acuzele de definire tautologică a constructelor analizate (de exemplu, inteligenţa este ceea ce testele de inteligenţă măsoară).

Caseta 1.8. *Formulări PICO pentru studii din zona psihologiei şi a psihoterapiei*

Studii clinice controlate	Studiul diferenţelor existente la nivelul depresiei (O) între adolescenţii delincvenţi (P) care au urmat o terapie cognitiv-comportamentală (I) şi cei care au beneficiat de un program de consiliere standard, uzual (C).
Studii longitudinale efectuate pe eşantioane largi	Copii adoptaţi (P) de către cupluri în care cel puţin unul dintre părinţi a suferit ulterior un episod psihopatologic din Axa I a DSM, necesitând tratament de specialitate (I) au un risc mai crescut să fie afectaţi de vreo tulburare mentală din Axa I (O) comparativ cu cei adoptaţi de cupluri fără istoric psihopatologic (C).
Studii corelaţionale cu rol predictiv	Introducerea evaluării conştiinciozităţii (I) la candidaţii pentru poziţii în domeniul vânzărilor din industria textilă (P) prezice performanţa lor ulterioară în muncă (O).

În aceeaşi direcţie a studiului unor probleme rezolvabile, o variantă utilă de lucru este de a formula tema studiată utilizând algoritmul PICO întâlnit în domeniul practicii validate ştiinţific. Formatul PICO (Spring, 2007) este utilizat pentru a formula întrebări de cercetare testabile. Denumirea reprezintă acronimul de la populaţia investigată (*Population*), intervenţia evaluată (*Intervention*), elementul comparativ evaluat (*Comparison*) şi de la rezultatul obţinut (*Outcome*). Dintre cele patru componente, elementul comparativ evaluat este singurul care poate avea un caracter opţional, însă doar pentru cazul acelor studii care pot fi derulate adecvat în absenţa unui grup de control (de exemplu, studiile corelaţionale).

1.5.4. Conectivitate cu teoriile şi studiile anterioare

Pentru a avea credibilitate ştiinţifică, orice teorie sau model nou propus ar trebui să poată explica datele obţinute, cel puţin la fel de bine ca teoriile deja existente. Mai mult, teoriile şi modelele noi devin superioare celor existente dacă reuşesc să explice dinamica rezultatelor acolo unde teoriile anterioare din domeniu au eşuat să facă acest lucru. În general, modelele explicative sunt interconectate între ele interdisciplinar. De pildă, modelul evoluţionist are aplicaţii variate în psihologie, biologie, economie. O teorie este cu atât mai solidă cu cât este susţinută de datele colectate din diverse domenii de studiu.

Aceste aspecte menţionate sunt importante de reţinut atunci când evaluăm potenţialul ştiinţific al unor modele alternative propuse. Astfel, Stanovich (2010) oferă exemplul a două intervenţii ipotetice menite să amelioreze dificultăţile de citire ale elevilor din primul ciclu şcolar, în contextul în care despre niciuna dintre ele nu există date empirice privind eficienţa lor terapeutică. Să denumim cele două forme de intervenţie A, respectiv B. Intervenţia A presupune antrenarea conştientizării din partea copilului a caracterului segmentar al limbajului la nivel fonologic (prin exerciţii de despărţire în silabe). Intervenţia B presupune antrenarea sensibilităţii la nivel vestibular a copiilor (prin exerciţii repetate de mers pe bârnă cu ochii închişi). În contextul în care intervenţia A se bazează pe unele dovezi empirice anterioare (de exemplu, existenţa unei corelaţii între dificultăţile de citire şi nivelul scăzut de conştientizare a caracterului segmentar al cuvintelor), iar intervenţia B nu se fundamentează pe rezultatele niciunui studiu anterior, este de preferat să se propună intervenţia A ca modalitate de lucru, chiar dacă încă nu sunt disponibile

rezultatele vreunui studiu clinic controlat care să demonstreze eficienţa intervenţiei de tip A.

În sfârşit, asociată ideii de conectivitate este şi principiul convergenţei dovezilor, conform căruia, ştiinţa progresează treptat, pe măsură ce anumite teorii primesc tot mai mult sprijin prin invalidarea explicaţiilor alternative. Niciun studiu nu este perfect. Orice studiu are o serie de limite privind validarea sa, ca urmare este nevoie de un set de dovezi convergente (unele experimentale, altele obţinute într-un context ecologic, altele longitudinale etc.) pentru a ajunge la un progres. De aceea, ştiinţa tinde să se bazeze din ce în ce mai mult pe rezultatele studiilor cantitative, de tip metaanalitic, pentru a răspunde unei probleme analizate şi pentru a găsi noi aspecte de investigat.

1.5.5. Scepticism şi căutarea explicaţiilor alternative

Ar putea părea neobişnuit, dar unul dintre aspectele importante într-un demers ştiinţific este tocmai atitudinea sceptică în faţa afirmaţiilor sau dovezilor prezentate de terţi. Astfel, chiar pentru situaţii destul de evidente precum acelea că frustrarea este asociată cu agresivitatea sau că nivelul de inteligenţă prezice performanţa şcolară, oamenii de ştiinţă au nevoie de dovezi concrete pentru că nu se pot baza doar pe simţul comun, pe intuiţie, sau pe opinia neverificată a vreunei autorităţi în domeniu.

Mai mult, chiar în contextul în care îi sunt prezentate dovezi în sprijinul afirmaţiilor respective, următorul pas într-un demers ştiinţific corect nu este acela de a jubila şi a proclama că acesta este adevărul absolut, ci de a pune alte întrebări cu rol cârcotaş precum: „Există alte studii care să contrazică rezultatele găsite de noi?” sau „Ce alte teorii

alternative (variabile confundate) ar putea explica (conduce la) acest rezultat?”.

Pentru a ilustra importanţa unei atitudini sceptice în ştiinţă voi apela la exemplul oferit de Coolican (2004), cu referire la războiul dintre giganţii băuturilor răcoritoare Coca-Cola şi Pepsi. Cei de la Pepsi au prezentat un studiu realizat pe băutorii de Coca-Cola conform căruia mai mult de jumătate dintre consumatorii obişnuiţi de Coca-Cola au considerat gustul băuturii Pepsi ca fiind mai bun decât cel al Coca-Cola, dacă băuturile erau servite din două pahare denumite M şi Q, fără a şti însă în care dintre pahare se servea Pepsi, respectiv Coca-Cola. În realitate, pentru o înregistrare cât mai corectă a studiilor, Pepsi era servită din paharul cu denumirea M, iar Coca-Cola din paharul cu denumirea Q. Impactul publicitar al acestui studiu a avut efect, raportul de vânzări în favoarea Coca-Cola scăzând de la un raport de 3 la 1, la unul de 2 la 1. Această evoluţie a vânzărilor i-a determinat pe cei de la Coca-Cola să argumenteze că studiul anterior a fost construit inadecvat, pentru că a oferit din start un avantaj pentru Pepsi Cola care fusese servit din pahare inscripţionate cu litera M, deoarece există o preferinţă implicită a oamenilor (cel puţin a vorbitorilor de limbă engleză) pentru litera M (de exemplu, mama) mai ridicată decât pentru litera Q. De aceea, au realizat un alt studiu în care participanţii erau manipulaţi să creadă că trebuie să aleagă între două băuturi răcoritoare, puse în pahare inscripţionate cu M, respectiv Q. S-a constatat că un număr semnificativ mai ridicat persoane preferau băutura din paharul M comparativ cu cea din paharul Q, deşi în amândouă era acelaşi conţinut – Coca-Cola! Deşi aparent, Coca-Cola a oferit un răspuns consistent, care a oferit o explicaţie alternativă pentru rezultatele primului studiu, alta decât cea dorită de cei de la firma Pepsi, niciunul dintre cele două studii nu a fost realizat adecvat pentru a răspunde la întrebarea

legată de preferinţa pentru una sau alta dintre băuturile răcoritoare. De exemplu, cel de-al doilea studiu, deşi a reuşit să ofere o explicaţie alternativă faţă de cea prezentată de cei de la Pepsi, nu a reuşit să indice dacă rezultatele din primului studiu se datorează în totalitate preferinţei implicite pentru anumite litere sau dacă acest efect este doar unul care moderează intensitatea preferinţei existente pentru Pepsi faţă de Coca-Cola, la nivel de gust.

În opinia noastră, cel puţin două măsuri de control a variabilelor străine ar fi trebuit luate în primul studiu pentru a elimina posibilele explicaţii alternative. În primul rând ar fi vorba de contrabalansarea tipului de băutură prezentată în paharele de tip M, respectiv de tip Q, precum şi de contrabalansarea ordinii de prezentare a acestor băuturi (pornind de la premisa că este posibil ca al doilea pahar gustat să nu aibă acelaşi efect gustativ decât primul, mai ales pentru participanţii însetaţi). Astfel, dacă jumătate dintre participanţi ar fi primit Pepsi în pahare de tip M, iar cealaltă jumătate în pahare de tip Q (şi viceversa pentru Coca-Cola) am fi controlat efectul preferinţei implicite pentru anumite litere şi l-am fi putut lua în calcul. De asemenea, dacă jumătate dintre participanţi ar fi băut mai întâi Pepsi, iar cealaltă jumătate ar fi băut mai întâi Coca-Cola, am fi putut controla şi lua în calcul şi efectul de ordine a degustării băuturilor. Un asemenea design, nu le-ar fi permis celor de la Coca-Cola să aducă drept explicaţie alternativă ideea că efectul se datorează preferinţei pentru anumite litere, nu diferenţelor de gust dintre băuturi. Desigur, chiar şi în cazul prezenţei contrabalansării, atitudinea sceptică faţă de rezultatele obţinute ar fi putut fi invocată. De pildă, s-ar putea ca rezultatele să fie distorsionate din cauza faptului că studiile nu au fost derulate în condiţii de tip dublu-orb. În cele două studii derulate, doar participanţii nu cunoşteau ce se află în cele două pahare, în schimb

experimentatorii ştiau. E posibil, chiar fără voia lor, să fi influenţat decizia participanţilor la studiu, cei de la Pepsi având o atitudine mai apropiată faţă de participanţi în momentul în care aceştia degustau din paharul M (de Pepsi), iar cei de la Coca-Cola să transmită involuntar mesaje non-verbale care să-i influenţeze în mai mare măsură pe participanţi să prefere băutura din paharul inscripţionat cu litera M. Un design dublu-orb, în care nici participanţii, nici asistenţii de cercetare care intrau direct în contact cu cei investigaţi, nu ar cunoaşte tipul de băutură servită, asociat cu aplicarea măsurilor de contrabalansare pentru a controla efectele de ordine sau alte variabile străine cum ar fi preferinţa pentru anumite litere, ar fi condus la un studiu experimental bine controlat, care să înlăture eventualele critici (explicaţii alternative) venite din partea celorlalţi.

Istoria ştiinţei este plină de asemenea exemple, poate cel mai celebru dintre ele fiind cel legat de malarie, a cărei denumire originară din limba franceză, provine tocmai de la credinţa falsă că era declanşată de aerul rău din zona mlaştinilor (Mitchell şi Jolley, 2006). Au fost efectuate chiar unele studii, arătând că există un număr mai mare de îmbolnăviri în locuinţele din apropierea bălţilor şi a mlaştinilor decât în zonele cu teren uscat. Desigur, ce s-a neglijat în aceste explicaţii iniţiale este faptul că şi ţânţarii, adevăratul mijloc de transmitere a bolii, se află într-un procent mai mare în zona bălţilor şi a mlaştinilor decât în zone cu teren arid.

Concluzii

Fascinaţia societăţii actuale pentru domeniul psihologiei este în creştere. În special mass-media preia şi dezbate teme din domeniul psihologiei într-o măsură mai mare decât se întâmplă în cazul altor ştiinţe. Însă deseori

dovezile ştiinţifice sunt înlocuite de simple speculaţii bazate pe intuiţie şi pe simţul comun. La aceasta se adaugă confuziile legate de asocierea psihologiei cu alte domenii precum parapsihologia, curentele de tip New Age, medicina alternativă, precum şi întreaga industrie de tip *self-help*. Însă, poate cel mai puternic pericol provine tocmai din tagma psihologilor. Destul de mulţi colegi nu reuşesc să facă distincţia dintre abordarea ştiinţifică şi pseudoştiinţă. Dacă la nivel stilistic, există un oarecare stereotip format despre pseudoştiinţe (vezi segmentul de jos al figurii 1.1), la nivelul conţinuturilor şi al metodelor utilizate diferenţa rămâne, pentru mulţi, una nesesizabilă. Pseudoştiinţele tind să fetişizeze rezultatele pozitive obţinute, să ignore dovezile contrare, să abuzeze de explicaţii *post hoc* fără a le supune verificării şi să-i acuze pe cei care contestă mesajul lor. Într-un fel, este naturală această tendinţă umană de a căuta dovezi pentru a susţine ceea ce crezi, cu atât mai mult într-un domeniu precum psihologia despre care multă lume are impresia că se pricepe. În ştiinţă se procedează într-un mod diferit. Ştiinţa nu este definită de tema pe care o studiază, ci de metodologia pe care o aplică în studierea temelor respective. Aceasta trebuie să fie riguroasă (empirism sistematic), iar rezultatele obţinute să se supună controlului public şi atitudinii sceptice manifestate de către colegii de breaslă. Cunoştinţele care trec cu succes prin aceste filtre sunt singurele care pot fi considerate ca fiind întemeiate, utile pentru evoluţia ştiinţei şi practicii psihologice.

Până la urmă este lăudabil faptul că cei mai mulţi dintre cei care aleg să studieze psihologia fac acest lucru pentru că „vor să-i ajute pe ceilalţi” sau pentru că „vor să se cunoască mai bine pe ei înşişi”. Însă, pentru a realiza aceste lucruri, este nevoie să se utilizeze informaţii şi tehnici validate empiric, pe care numai o abordare ştiinţifică le poate oferi. O asemenea abordare merită promovată inclusiv în rândul acelora care se

simt atraşi de misterele psihologiei, care utilizează teorii şi concepte psihologice doar pentru a da sens şi semnificaţie propriilor experienţe de viaţă şi care sunt mai mereu predispuşi să ofere *post hoc* interpretări psihologizante pentru evenimentele trăite. Pentru mulţi dintre ei poate părea straniu să abordezi ştiinţific teme precum dragostea romantică, însă înţelegerea mecanismelor psihologice implicate în acest caz, într-o manieră validată ştiinţific, va ajuta la prevenirea şi la combaterea depresiei, a violenţei domestice şi va contribui la ameliorarea vieţii de cuplu. Ştiinţa psihologiei nu face altceva decât să filtreze informaţiile din zona psihologiei pe baza datelor factuale şi să corecteze credinţele greşite, astfel încât să-i ajute mai eficient pe oameni şi să contribuie, prin cunoaştere, la progresul societăţii.

Lecturi suplimentare

Skeptical Inquirer (revistă de specialitate)

Această revistă este editată cu scopul de a evalua într-o manieră critică, folosind metode ştiinţifice, orice afirmaţie sau concluzie exagerată (neacoperită), care a trecut de filtrul evaluării colegiale şi a fost publicată în reviste ştiinţifice recunoscute. Tema preferată de analiză este domeniul parapsihologiei, dar fără a se limita doar la aceasta.

Scott O. Lilienfeld, Steven Jay Lynn, John Ruscio, Barry L. Beyerstein (2010). *50 de mari mituri ale psihologiei populare. Înlăturarea concepţiilor greşite despre comportamentul uman*. Bucureşti: Editura Trei.

Cartea prezintă într-o manieră jurnalistică 50 de mituri celebre din domeniul psihologiei (de exemplu, „ne folosim doar 10% din capacitatea creierului") pe care le analizează din perspectiva gradului lor de fundamentare ştiinţifică. Subtitlul cărţii este unul ilustrativ: înlăturarea concepţiilor greşite despre comportamentul uman.

Scott O. Lilienfeld, Steven Jay Lynn, Jeffrey M. Lohr (2004). *Science and Pseudoscience in Clinical Psychology*. New York: Guilford Press.

Lucrarea este principala referinţă bibliografică care delimitează între abordările ştiinţifice şi cele pseudoştiinţifice în domeniul psihoterapiei. În linii mari lucrarea analizează cele mai întâlnite

forme de intervenţie psihoterapeutică, precum şi practicile de evaluare clinică, din perspectiva măsurii în care acestea îndeplinesc criteriile pentru a fi considerate demersuri validate ştiinţific.

Daniel David (2003). *Castele de nisip. Ştiinţă şi pseudoştiinţă în psihologia clinică*. Bucureşti: Editura Tritonic.

Cartea constituie primul semnal de alarmă în spaţiul românesc cu privire la răspândirea curentelor pseudoştiinţifice în domeniul psihologiei clinice şi a psihoterapiei.

Keith E. Stanovich (2010), *How to think straight about psychology* (ediţia a IX-a). Boston: Pearson.

Lucrarea este una de referinţă prin faptul că reuşeşte să prezinte publicului, într-un limbaj accesibil, cum să diferenţieze între abordările riguroase, ştiinţifice din psihologie şi încercările pseudoştiinţifice din domeniu. Este o lucrare care ar trebui introdusă obligatoriu în pregătirea academică a oricărui student la psihologie.

Bibliografie

Academia Premiilor Nobel (2012). The Oficial Web Site for Nobel Prize. Wilhelm Wundt. Găsit la adresa http://www.nobelprize.org/

APA (2009). Presidential Task Force on the Future of Psychology Practice Final Report. Găsit la adresa http://www.apa.org/pubs/info/reports/future-practice.pdf.

Bayerstein, B.L. (1996). Distinguishing science from pseudoscience. Găsit la adresa: http://www.sld.cu/galerias/pdf/sitios/revsalud/beyerstein_cience_vs_pseudoscience.pdf

Bem, D.J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, pp. 404-425.

Bringmann, W.G., Ungerer, G.A., Bringmann, M.W. (1995). Wilhelm Wundt and the workers' educational movement. *Psychologie und Geschichte*, 6, pp.233-243.

Callahan, R. J. (2001). Thought Field Therapy: Response to our critics and a scrutiny of some old ideas of science. *Journal of Clinical Psychology*, 57, pp. 1251-1260.

Coolican, H. (2004). *Research methods and statistics in psychology* (ediţia a IV-a). Londra: Hodder & Stoughton.

David, D. (2010). *Scrisoare către babilonieni*. Disponibilă la adresa http://www.psihologiaonline.ro/component/content/article/471-psihologul-daniel-david-avertizeaza-politicienii.html

FDSC (2010). Newly emerging needs of children in Romania. An exploratory study. raport al Fundaţiei pentru Dezvoltarea Societăţii Civile, găsit la adresa http://www.scribd.com/doc/53063540/Newly-Emerging-Needs-of-Children-in-Romania-Exploratory-Study

Goodheart, C.D. (2010). Economics and psychology practice: What we need to know. *Professional Psychology: Research and Practice*, 41, pp. 189-195.

Gravetter, F.J., Forzano, L.-A. (2009). *Research methods for the behavioral sciences* (ediţia a IV-a). Belmont: Wadsworth.

Henriques, G.R., Sternberg, R. (2004). Unified professional psychology: Implications for the combined-integrated model of doctoral training. *Journal of Clinical Psychology*, 60, pp. 1051-1063.

Herbert, J.D., Lilienfeld, S.O., Lohr, J.M., Montgomery, R.W., O'Donohue, W.T., Rosen, G.M., Tolin, D.F. (2000). Science and pseudoscience in the development of eye movement desensitization and reprocessing: Implications for clinical psychology. *Clinical Psychology Review*, 20, pp. 945-971.

Hyman, R. (1977). Cold reading: How to convince strangers that you know all about them. *The Skeptical Inquirer*, 1, pp. 18-37.

Jansz, J. (2004). Psychology and society: an overview. În J. Jansz şi P. van Drunen (eds.), *A social history of psychology* (pp. 12-44). Oxford, UK: Blackwell Publishing.

Kaminer, W. (1992). *I'm Dysfunctional, You're Dysfunctional: The Recovery Movement and Other Self-Help*. New York: Addison Wesley.

Kendler, H.H. (1999). The role of value in the world of psychology. *American Psychologist*, 54, pp. 828-835.

Knapp, S., Keller, P.A. (2001). Professional associations' strategies for revitalizing professional psychology. *Professional Psychology: Research and Practice*, 32, pp. 71-78.

Levant, R.F. (2003, Ianuarie 1). Making psychology a household word. *National Psychologist*, găsit la adresa http://nationalpsychologist.com/2003/01/making-psychology-a-household-word/10631.html

Lilienfeld, S.O., Landfield, K. (2008). Science and pseudoscience in law enforcement: A user-friendly primer. *Criminal Justice and Behavior*, 35, pp. 1215-1230.

Lilienfeld, S.O., Lohr, M., Morier, D. (2001). The teaching of courses in the science and pseudoscience of psychology. *Teaching of Psychology*, 28, pp. 182-191.

Lilienfeld, S.O., Lynn, S.J., Lohr, J.M. (2004). Science and pseudoscience in clinical psychology: Initial thoughts, reflections, and considerations. În S.O. Lilienfeld, S.J. Lynn şi J.M. Lohr (eds.), *Science and pseudoscience in clinical psychology* (pp. 1-14). New York: Guilford.

Lilienfeld, S.O., Lynn, S.J., Ruscio, J., Beyerstein, B.L. (2010). *50 de mari mituri ale psihologiei populare*. Bucureşti: Editura Trei.

Louw, J. (2002). Psychology, history, and society. *South-African Journal of Psychology*, 32, pp. 1-8.

Marin, G. (2009). Ce este eneagrama?. Găsit la adresa http://www.empower.ro/inteligenta-emotionala/ce-este-eneagrama/

Mitchell, M.L.. Jolley, J.M. (2006). *Research design explained* (ediţia a VI-a). Belmont :Wadsworth Publishing.

Morar, O. (2005). Intelectualii români şi „chestia evreiască”. *Contemporanul. Ideea Europeană*, 6, disponibil la adresa http://www.romaniaculturala.ro/articol.php?cod=7466

New World Encyclopedia (2008). Wundt Wilhelm. Disponibil la adresa http://www.newworldencyclopedia.org/entry/Wilhelm_Wundt

Patrick, J. (2005). The economic value of psychology in Australia: 2001. *Australian Psychologist*, 40, pp. 149-158.

Pelham, B.W., Blanton, H. (2007). *Conducting research in psychology. Measuring the weight of smoke* (ediţia a III-a). Belmont: Thompson Wadsworth.

Popper, K. (1981). *Logica cercetării*. Bucureşti: Editura Ştiinţifică.

Postman, N. (1986). *Amusing ourselves to death: Public discourse in the age of show business*. New York: Penguin.

Saeman, H. (1998, July). Survey: Psychologists’ incomes plummet. *The National Psychologist*, 7, pp. 1-2.

Savage, M. (2009). Psychology and contemporary society. *Modern Intellectual History*, 6, pp. 627-636.

Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. *Journal of Clinical Psychology*, 63, pp. 611-631.

Stanovich, K.E. (2010). *How to think straight about psychology* (ediţia a IX-a). New York: Pearson.

Vlăsceanu, L., Duşa, A., Rughiniş, C. (coord.) (2010). Ştiinţă şi societate. Interese şi percepţii ale publicului privind cercetarea ştiinţifică şi rezultatele cercetării. Raport de cercetare disponibil la adresa http://www.stisoc.ro/docs/STISOC2010_Raport%20de%20cercetare.pdf.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (ediţia a IV-a). San Antonio: Pearson.

Wikipedia (2009). Bruno Bettelheim. Găsit la adresa http://en.wikipedia.org/wiki/Bruno_Bettelheim.

1. Deşi cererea de servicii psihologice este în creştere la nivel mondial, trebuie menţionat faptul că ritmul de creştere al cererii este sub cel de creştere a ofertei de servicii psihologice, fapt reflectat prin tendinţa de scădere a nivelului de salarizare pentru psihologi (Saeman, 1998), de creştere a şomajului în rândul tinerilor absolvenţi de psihologie (Knapp şi Keller, 2001) şi de creştere a numărului de psihologi care au o opinie defavorabilă despre această profesie şi care i-au în considerare o schimbare a domeniului de activitate (Knapp şi Keller, 1998 *apud* Henriques şi Sternberg, 2004).

2. „In mathematical terms, Derrida’s observation relates to the invariance of the Einstein field equation… under nonlinear space-time diffeomorphisms (self-mappings of the space-time manifold that are infinitely differentiable but not necessarily analytic)… In this way, the infinite-dimensional invariance group erodes the distinction between observer and observed; the p of Euclid and *G* of Newton, formally thought to be constant and universal, are now perceived in their ineluctable historicity; and the putative observer becomes fatally decentered,

disconnected from any epistemic link to a space-time point that can no longer be defined by geometry alone.”

3. În aceeaşi direcţie, aspecte care ţin direct de sfera spiritualităţii pot fi abordate ştiinţific (de exemplu, influenţa rugăciunilor asupra emoţiilor, substratul neurobiologic al stărilor meditative, rolul religiozităţii în sănătatea mentală etc.).

4. Spre deosebire de alte caracteristici ale unui abordări ştiinţifice (vezi figura 1.2), empirismul sistematic are o sferă de aplicabilitate limitată, îndeosebi la nivelul ştiinţele empirice (naturale, ale vieţii, sociale), fiind mai puţin potrivit pentru a descrie situaţia întâlnită în cadrul ştiinţelor formale (matematică, logică). Ştiinţele formale (analitice) nu sunt interesate de validarea teoriilor pe baza observaţiilor colectate, ci de analiza proprietăţilor unor sisteme formale pe baza unor definiţii şi a unor reguli de operare.

5. Unii autori consideră drept demersuri ştiinţifice şi topicile care nu reprezintă probleme rezolvabile în prezent (de exemplu, căutarea fiinţelor extraterestre), atâta timp cât metodologia din studiile respective respectă condiţia falsificabilităţii. În asemenea cazuri nu se oferă însă răspunsuri definitive cu privire la tematica investigată, iar uneori se poate ajunge la situaţia stranie în care lipseşte însăşi dovada existenţei obiectului de studiu. Ceea ce contează este ca problemele să fie rezolvabile în principiu, nu neapărat în prezent. Astfel pot rezulta ipoteze interesante şi se poate acumula un corp important de cunoştinţe (de exemplu, studiul condiţiilor extreme în care poate apărea viaţa).

Capitolul 2
Practica psihologică validată ştiinţific

2.1. Definirea şi descrierea practicii validate ştiinţific

Capitolul anterior are meritul de a trece în revistă principalele diferenţe dintre o abordare ştiinţifică şi una pseudoştiinţifică, cu referire îndeosebi la specificul psihologiei ca ştiinţă. Însă, abordarea ştiinţifică nu vizează doar sfera academică, ci ar trebui să se reflecte inclusiv în practica psihologilor. De altfel, un deziderat frecvent întâlnit în programele de formare profesională a psihologilor este acela de a forma ceea ce în limba engleză se numeşte *scientist practitioners*, cu alte cuvinte, de a pregăti specialişti care să fie buni practicieni, dar care să posede şi abilităţi de cercetare. Idealul unui „practician-cercetător" este acela de a-şi fundamenta practica pe soluţii validate ştiinţific. Pentru a năzui spre acest deziderat, practicienii trebuie să fie consumatori critici ai literaturii de specialitate. Aceasta presupune atât căutarea de informaţii ştiinţifice pertinente pentru problemele din practica profesională, cât şi capacitatea de a discerne între diferitele soluţii de intervenţie existente, luând în calcul eficienţa lor, dovedită prin studii empirice. De asemenea, un practician-cercetător manifestă un interes direct pentru producerea de cunoaştere, reflectată printr-un comportament activ în manifestări

profesionale (Gaudiano şi Statler, 2001; Overholser, 2010). Această viziune despre practica psihologică a fost teoretizată iniţial încă din anii ’50 ai secolului trecut, fiind cunoscută şi sub numele de modelul Boulder, un model originar din zona psihologiei clinice (Striker, 1997).

Un alt concept important este cel de practică validată ştiinţific (*evidence-based practice* sau EBP)[1]. Între conceptul de EBP şi cel de practician-cercetător există o suprapunere ridicată, chiar dacă cele două aspecte nu au un istoric comun. Practica validată ştiinţific (EPB) îşi are originea în domeniul ştiinţelor medicale, unde a cunoscut o dezvoltare foarte puternică. Atât ca urmare a unor similitudini ridicate dintre domeniul clinic al medicinei şi al psihologiei, cât şi ca urmare a similarităţilor conceptuale dintre conceptul de practician-cercetător şi cel de specialist care oferă servicii validate ştiinţific (bazate pe dovezi empirice), transferul dintre medicină spre psihologie s-a făcut mai întâi în zona psihologiei clinice, dar se extinde din ce în ce mai mult în toate ariile de interes ale psihologiei. Practica validată ştiinţific vizează în primul rând cerinţa ca în deciziile din practica profesională să se ia în calcul implementarea de soluţii pentru care există dovezi validate ştiinţific cu privire la eficienţa lor. Deşi se are în vedere rolul de producător de cunoaştere (realizator de cercetări), accentul cade pe rolul de consumator critic al literaturii de specialitate, în vederea identificării soluţiilor cu cel mai ridicat nivel de eficienţă. La fel de importantă este expertiza profesională a practicianului de a putea decide contextual dacă soluţiile oferite de literatura de specialitate sunt potrivite pentru problema concretă de rezolvat. Astfel, pe lângă aspectele ştiinţifice, se ţine seama de nevoile şi aşteptările clientului, astfel încât decizia de a selecta o anumită formă de intervenţie să se facă având consimţământul clientului.

Aşa cum am precizat, cele două concepte discutate diferă ca origine. Dacă conceptul de practician-cercetător a fost dezvoltat de psihologii clinicieni, ideea practicii validate ştiinţific îşi are originea în domeniul medicinei. De altfel, o definiţie consacrată a EBP susţine că aceasta este un proces care implică analizarea într-o manieră responsabilă, explicită, a dovezilor existente în literatura de specialitate cu privire la eficienţa diferitelor forme de intervenţie, cu scopul de a lua decizii adecvate cu privire la modul de îngrijire a unui pacient (Sackett, Rosenberg, Gray, Haynes şi Richardson, 1996).

În sfârşit, în literatura de specialitate unele surse (APA, 2006) fac o altă delimitare între EBP şi intervenţiile validate ştiinţific (*empirically supported treatments* sau EST). Conform acestora, EBP, deşi face apel la identificarea intervenţiilor validate ştiinţific pentru problema avută, presupune mai mult decât atât, fiind un demers complex de tip ideografic (APA, 2006), în timp ce intervenţiile validate ştiinţific constituie o componentă integrată în EBP, rezultată în urma unei abordări de tip nomotetic (de exemplu, analiza studiilor de specialitate pentru a răspunde la întrebarea: „este terapia familial-sistemică o modalitate eficientă de intervenţie în cazul adolescenţilor care au aflat recent că au fost înfiaţi?”). Astfel, EST vizează cu precădere intervenţia, în timp ce EBP vizează cu precădere clientul, utilizând informaţiile cu privire la intervenţiile validate ştiinţific ca unul dintre inputurile necesare pentru a lua cea mai bună decizie pentru clientul asistat (APA, 2006, Levant şi Hasan, 2008).

Astfel, conform acestei viziuni EBP se referă la procesul integrării informaţiilor din literatura de specialitate (cu privire la cele mai eficiente intervenţii validate ştiinţific), cu expertiza practică, în contextul analizei preferinţelor şi a caracteristicilor culturale şi demografice ale clienţilor care solicită servicii psihologice (APA 2006, Spring, 2007), pentru a se

atinge efectul dorit de client (de exemplu, reducerea nivelului de conflict muncă-familie, creşterea satisfacţiei maritale, reducerea anxietăţii sociale, creşterea performanţei şcolare etc.). Acest model de conceptualizare a EBP este cunoscut şi sub numele de modelul scaunului cu trei picioare (*the three legged-stool*) (Spring, 2007).

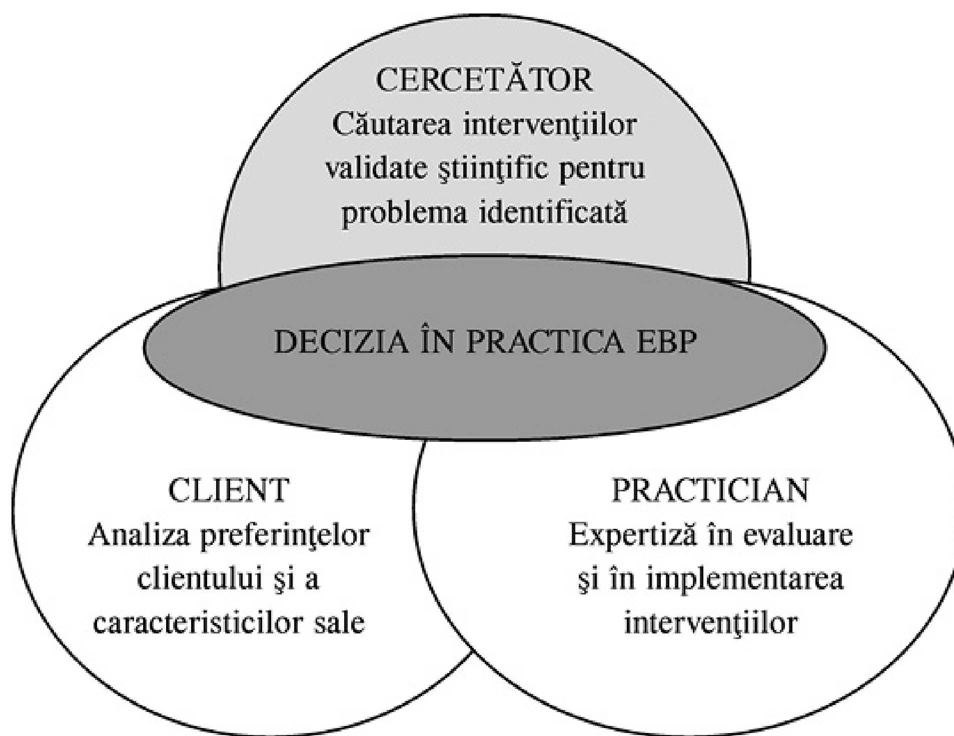


Figura 2.1. *Cele trei braţe ale practicii validate ştiinţific*

Pe scurt, EBP înseamnă a lua decizii în baza expertizei profesionale, a caracteristicilor clientului şi a informaţiilor la zi cu privire la cele mai eficiente tipuri de intervenţie pentru a soluţiona problema întâlnită. Conceptul de EBP s-a răspândit rapid din domeniul medicinei spre alte profesii din zona sănătăţii (psihologie clinică şi psihoterapie, asistenţă medicală, farmacie), dar şi în domenii diferite precum management, pedagogie sau asistenţă socială (Briner şi Rousseau, 2011; Craig şi Smyth, 2002; Gibbs, 2002; Kratochwill şi Shernoff, 2004; Noonan, 2009;

Norcross, Beutler şi Levant, 2006; Rousseau, 2006; Sackett, Strauss, Richardson, Rosenberg şi Haynes, 2000 etc.).

Pentru a vedea mai concret ce ar însemna o abordare EBP în practica psihologică, vom ilustra în tabelul 2.1 întrebările de bază la care ar trebui să răspundă un psiholog organizaţional atunci când ar fi solicitat, în calitate de consultant, să rezolve problema fluctuaţiei de personal dintr-o companie. În funcţie de răspunsurile găsite la aceste întrebări (sau măcar la cea mai mare parte a acestora), psihologul va căuta să ia cea mai bună decizie pentru a soluţiona problema respectivă.

Tabelul 2.1. *Întrebări specifice unei abordări EBP pentru reducerea fluctuaţiei de personal*

Perspectiva practicianului	A1. Există cazuri similare anterioare în portofoliul meu de servicii? Ce soluţii au fost propuse? Au funcţionat? A2. Care cred că sunt cauzele şi soluţiile posibile în acest caz? A3. Este o situaţie specifică companiei sau este afectată întreaga industrie de ramură şi/sau zonă în care este localizată firma? A4. Care este profilul organizaţiei beneficiare? A5. Care este politica actuală a companiei pentru a preveni fluctuaţia de personal? A6. Există date interne utile (de exemplu, chestionare aplicate la plecarea din firmă pentru a vedea motivaţiile specificate/profilul celor care pleacă sau alte studii care să indice situaţia actuală privind intenţia de a părăsi organizaţia)? A7. Care sunt cauzele fluctuaţiei în opinia conducerii companiei? A8. Ce soluţii au aplicat competitorii direcţi pentru a rezolva această problemă?
Perspectiva cercetătorului	B1. Care este rata fluctuaţiei de personal în acest domeniu? B2. Care sunt principalele cauze (factori de risc) ale fluctuaţiei de personal, rezultate din analiza critică a literaturii de specialitate? B3. Există intervenţii validate ştiinţific ţintite pentru a reduce fluctuaţia de personal? B4. Care este utilitatea lor practică (mărimea efectului)? B5. Cât de puternice sunt dovezile pe care le avem despre eficienţa intervenţiilor identificate (care este nivelul lor de validare ştiinţifică, de suport empiric)? B6. Cât de relevante sunt studiile respective (cât de bine se potrivesc) luând în calcul contextul particular al organizaţiei beneficiare?

	B7. Care dintre intervenţiile eficiente s-a dovedit a avea cel mai bun raport cost-beneficiu?
Perspectiva clientului	C1. Soluţia propusă ar stârni reactanţă (rezistenţă) din partea personalului firmei? C2. Care este opinia conducerii despre soluţia propusă? C3. Ce alte soluţii ar fi preferate/sunt luate în calcul de către conducerea companiei? C4. Ce consecinţe negative (efecte secundare nedorite) ar putea apărea în urma aplicării intervenţiei propuse? C5. Care ar fi poziţia conducerii faţă de aceste consecinţe secundare? C6. Ce dificultăţi sunt probabile să apară în procesul de implementare a intervenţiei? C7. Ce alte probleme neidentificate explicit de client pot fi rezolvate (diminuate) prin intervenţia propusă?

Totuşi, o serie de surse recente tind să trateze cele două concepte de EST şi EBP ca fiind sinonime (APA, Divizia 12 de Psihologie Clinică, 2012). Conform acestora, EBP este un concept integrator care indică orice tip de activitate (intervenţii validate ştiinţific – EST, evaluări validate ştiinţific – ESA, *evidence based assessment*) pentru care există suport ştiinţific. Conform acestei interpretări, EBP nu reprezintă altceva decât EST, ci reflectă un concept integrator/sinonim, pentru a cuprinde orice activitate practică (inclusiv componenta de evaluare psihologică), organizată pe baza aceloraşi principii enunţate mai sus. De aceea, destul de mulţi autori utilizează EBP şi EST alternativ, fără a face distincţie între ei (David, comunicare personală, 22 aprilie 2012).

2.2. Necesitatea tranziţiei către o practică validată ştiinţific

Cadrul conceptual privind EBP este unul extrem de atrăgător, fapt ce a condus la expansiunea şi îmbrăţişarea modelului dincolo de zona ştiinţelor medicale, unde rămâne însă cel mai bine răspândit. În realitate, însă, cei mai mulţi practicieni, inclusiv din domeniu medicinei, se raportează la acest cadru ca la un model normativ, prescriptiv. Dacă analizăm nivelul descriptiv, vom constata faptul că cei mai mulţi psihologi practicieni rămân blocaţi la nivelul perspectivei practicianului din tabelul 2.1.

Din această perspectivă, în cazul EBP există o mare discrepanţă dintre teorie şi practică. De pildă, doar aproximativ jumătate dintre medici studiază literatura de specialitate pentru a lua decizii în practica lor curentă (Bennett *et al.*, 2003; Humphris, Littlejohns, Victor, O'Halloran şi Peacock, 2000). Nu se cunoaşte însă cifra celor care evaluează critic informaţia citită, nici a celor care decid o soluţie luând în calcul şi preferinţele/valorile pacientului. De asemenea, mai puţin de 1% dintre manageri citesc studii de specialitate pentru a-şi fundamenta deciziile (Rynes, Giluk şi Brown, 2007). Discrepanţa dintre modelul teoretic şi proba realităţii este una foarte mare şi în domeniul psihologiei (Wilson, Armoutliev, Yakunina şi Werth Jr., 2009).

Din nefericire, cel mai adesea deciziile practicienilor se bazează mai degrabă pe cutume profesionale introduse, fie în baza unei analize raţionale, a intuiţiei sau a simţului comun, fie pur şi simplu în baza unui comportament mimetic („pentru că aşa au procedat şi alţii”). Foarte rar deciziile din practica psihologică se bazează pe dovezi empirice, validate ştiinţific.

Caseta 2.1. *Două cazuri specifice de intervenţii nevalidate ştiinţific*

Exemplul 1. *Programul de „înfricoşare directă” (*scared straight*), adaptare după Petrosino, Turbin-Petrosino şi Buehler (2009)*

Ideea programului a venit de la un film documentar distins cu premii de specialitate în 1978, în care un grup de adolescenţi delincvenţi era pus faţă în faţă cu un grup de infractori periculoşi. Aceştia îi intimidau pe adolescenţi cu scopul de a-i speria şi de a-i determina să realizeze că se află pe un drum greşit care le va aduce multă suferinţă, dacă nu renunţau la comportamentele lor delincvente.
Ideea a fost preluată şi introdusă în practică pe scară largă, îndeosebi în SUA şi în Marea Britanie pe deţinuţi adolescenţi şi tineri. Aceştia erau expuşi în direct în faţa unor confraţi condamnaţi la închisoare pe viaţă cu scopul de a conştientiza consecinţele unei asemenea perspective şi a-i speria să nu ajungă ca ei, astfel încât să nu comită acte de recidivă. Asemenea programe par să funcţioneze deoarece 94% dintre cei care au intrat în program nu au recidivat ulterior (comparaţii făcute exclusiv pe grupul beneficiar, fără vreun grup de control).
Surpriză însă! În urma unei analize adecvate (ce presupune compararea evoluţia deţinuţilor care au beneficiat de acest program cu cele ale deţinuţilor din grupurile de control, în cadrul unor studii randomizate), se observă un efect dăunător al programului. Şansele de a comite infracţiuni sunt cu aproximativ 70% mai mari în cazul celor care au participat la acest program decât în cazul celor care nu au „beneficiat” de o asemenea măsură corectiv-educativă. Cu excepţia unui singur studiu, în toate celelalte cercetări au fost obţinute efecte negative semnificative statistic.

Exemplul 2. *Programele educative adresate tinerilor şoferi din Marea Britanie cu scopul de a reduce numărul de accidente rutiere, adaptare după Ig şi Kwan (2001), actualizat în 2008*

Un asemenea program educativ ar fi necesar dacă am lua în considerare două cifre statistice: tinerii cu vârsta între 17 şi 21 de ani reprezintă 7% din populaţia celor care deţin un permis de conducere în UK, dar produc 13% din totalul accidentelor, ceea ce înseamnă că este un grup-ţintă predispus să comită accidente de circulaţie. În baza unui sondaj de opinie (chestionar) cu o rată de răspuns de 36%, s-a constatat că tinerii care au urmat un program de educaţie rutieră şi-au modificat atitudinea în direcţia oferirii unor răspunsuri ce indică o atitudine prudentă în trafic, reducând riscul apariţiei accidentelor de circulaţie. Această comparaţie s-a bazat pe un studiu de tip pretest-posttest fără un grup de control, cu o distanţă de 3 luni între testări, acestea având loc înainte şi după implementarea programului educativ.
Statul britanic a investit mulţi bani în extinderea acestui tip de programe pe scală largă, deşi datele obţinute în urma unei evaluări serioase a literaturii de specialitate evidenţiază faptul că asistăm la o risipă de bani publici. Astfel, analizând eficienţa acestor programe de formare (incluzând doar studii randomizate) se constată că nu există diferenţe

semnificative între cei care urmează asemenea programe şi cei din grupul de control cu privire la frecvenţa implicării în accidente de circulaţie. Cu alte cuvinte, nu există sprijin empiric solid pentru eficienţa acestor programe de educaţie rutieră.

Dincolo de specificul celor două problematici discutate, ar fi esenţial pentru orice practician din sfera psihologiei, care adoptă o atitudine profesionistă, să înţeleagă că:

a) pentru rezolvarea diverselor probleme apărute nu este suficient ca intervenţiile selectate să se bazeze doar pe opinii ale specialiştilor cu privire la ceea ce ar putea să funcţioneze; de exemplu, a afirma că statul ar trebui să investească mai mulţi bani în programe de educaţie rutieră pentru a reduce numărul tinerilor implicaţi în accidente de circulaţie pare o soluţie de bun-simţ, întâlnită frecvent în discursul public sau de tip jurnalistic; este însă nevoie să existe dovezi empirice că programele respective ar fi eficiente pentru a lua în calcul asemenea soluţii într-o abordare de tip EBP;
b) dovezile empirice în favoarea uneia sau alteia dintre formele de intervenţie nu sunt egale din perspectiva validităţii lor ştiinţifice (de exemplu, studiile randomizate sunt preferabile studiilor de tip non-experimental); de aceea adepţii EBP ar trebui să poată distinge între diferite tipuri de studii, din perspectiva calităţii dovezilor oferite de acestea (vezi detalii în secţiunea 2.3.2 a acestui capitol);
c) la ora actuală nu există dovezi validate ştiinţific pentru orice tip de intervenţie propusă; de aceea ar trebui să se facă o distincţie între intervenţiile derulate în absenţa dovezilor validate ştiinţific (pentru care nu există studii) şi intervenţiile care nu sunt sprijinite de dovezile colectate (studiile nu susţin eficienţa intervenţiei); intervenţiile din cea de-a doua categorie ar trebui eliminate din practica curentă, în timp ce intervenţiile din prima categorie pot fi luate în seamă doar dacă nu

există alte intervenţii validate ştiinţific, a căror eficienţă a fost demonstrată;

d) din perspectiva EBP, identificarea unei intervenţii validate ştiinţific în urma analizei critice a literaturii de specialitate este o măsură necesară, chiar esenţială, dar nu şi suficientă; alegerea intervenţiei ar trebui să ţină seama şi de alte aspecte existente precum: (1) resursele avute la dispoziţie (raportul cost-eficienţă, vezi capitolul 6); (2) validitatea ştiinţifică a cadrului conceptual, a teoriei pe care se bazează intervenţia (detalii în secţiunea 2.3.3 a acestui capitol); (3) caracteristicile individului sau a grupului pe care se doreşte a fi implementată soluţia.

Pentru explicaţii suplimentare legate de ultimul punct menţionat (d), voi ilustra cazul abandonului şcolar. Impactul pe termen mediu şi lung al acestei probleme este unul negativ, atât pentru indivizii în cauză (şomaj crescut, risc crescut de infracţionalitate, calificare profesională scăzută), cât şi pentru societate, prin costurile sociale pe care le generează.

În vederea prevenirii abandonului şcolar au fost încercate de-a lungul timpului mai multe tipuri de intervenţii care ar putea fi grupate în: (a) intervenţii educaţionale la nivel şcolar (restructurarea curriculei, a orarului pentru activităţile de şcoală/reducerea raportului de elevi per cadru didactic etc.); (b) intervenţii educaţionale la nivel individual (meditaţii, asistenţă în rezolvarea temelor de casă); (c) intervenţii educative în comunitate (implicarea elevilor în activităţi educative non-formale cum ar fi proiecte comunitare de asistenţă a vârstnicilor pentru a surprinde elemente ce ţin de biologie, psihologie, istorie etc.); (d) intervenţii comunitare legislative (de exemplu, programe de tip „cornul şi laptele” sau amendarea părinţilor care nu îşi monitorizează copiii şi care

nu se îngrijesc ca aceştia să frecventeze şcoala în ciuda unor atenţionări repetate prealabile primite prin scrisori de informare); (e) programe de consiliere a părinţilor cu privire la importanţa şcolii şi la consecinţele abandonului şcolar; (f) intervenţii cu scopul dezvoltării unor abilităţi psihologice individuale la elevi (îmbunătăţirea nivelului de autoeficacitate educaţională, reducerea distresului legat de şcoală, îmbunătăţirea atitudinii faţă de şcoală) etc.

În faţa unui număr atât de însemnat de opţiuni, primul impuls ar fi acela de a decide intervenţia cea mai potrivită în urma unei dezbateri între specialişti, luând în calcul opiniile acestora sau experienţa unor studii de caz.

Într-o perspectivă EBP, demersul de soluţionare este diferit, modelul standard impunând implementarea succesivă a celor cinci paşi necesari (Sackett *et al.*, 2000):

(1) reformularea problemei principale sub forma unei întrebări de investigat (pe cât posibil sub forma modelului PICO explicitat în cadrul secţiunilor 1.5.3 şi 7.2.1 din carte); în exemplul oferit, să spunem că se urmăreşte implementarea unei soluţii la nivel local (şcolile arondate psihologului şcolar), motiv pentru care soluţiile (a) şi mai ales (d) nu sunt fezabile, depinzând de modificări la nivel de sistem, iar soluţia (e) este puţin aplicabilă, deoarece este vorba de o comunitate cu foarte mulţi părinţi plecaţi la muncă în străinătate; în consecinţă rămân în analiză pentru pasul 2 soluţiile (b), (c) şi (f) şi întrebarea de a vedea care dintre cele trei variante este mai eficientă în prevenirea abandonului şcolar;

(2) analiza literaturii de specialitate în vederea identificării de dovezi empirice valide din punct de vedere ştiinţific, cu privire la cea mai eficientă formă de intervenţie pentru a preveni abandonul şcolar,

precum şi a factorilor moderatori care pot augmenta efectul (vârsta şcolarilor, tipul de comunitate – urban/rural etc.); în această direcţie sunt căutate în special studii metaanalitice (vezi capitolul 7); pentru problematica abandonului este identificat un astfel de studiu în cadrul colecţiei Campbell (Wilson, Tanner-Smith, Lipsey, Steinka-Fry şi Morrison, 2011) ce sintetizează 167 de studii; concluzia studiului metaanalitic este aceea că toate cele trei soluţii sunt eficiente, reducând abandonul şcolar (vezi figura 2.2);

Tabelul 2.2. *Trei criterii şi întrebări fundamentale aferente întâlnite în faza de evaluare critică a analizei de specialitate (pasul 3)*

Validitatea	Mă pot încrede în rezultatele studiilor respective?
Importanţa practică	Efectul produs este suficient de mare pentru a avea o relevanţă practică?
Aplicabilitate	Pot utiliza aceste informaţii în cazul concret de rezolvat?

(3) evaluarea critică a rezultatelor provenite din analiza literaturii de specialitate; deşi toate cele trei soluţii par a fi eficiente în prevenirea ratei de abandon şcolar, cea care implică elevii în programe educaţionale non-formale ce presupun servicii pentru comunitate pare a fi cea mai eficientă (riscul de abandon şcolar fiind de 3,53 ori mai mic pentru cei care au fost implicaţi în astfel de programe comparativ cu cei aflaţi în loturi de control); totuşi, pentru acelaşi tip de intervenţie se observă şi variabilitatea cea mai mare în eficienţa programelor (intervalul de încredere sugerând valori cuprinse între 1,90 şi 6,54), ceea ce înseamnă că efectul este dependent de o multitudine de factori, fiind mai puţin robust; tot din perspectiva încrederii în rezultate se poate urmări dacă aceste valori se bazează preponderent pe studii experimentale (randomizate) sau sunt obţinute în special din studii de tip cvasiexperimental (vezi detalii despre acest

subiect în secţiunea 2.3.2 şi în capitolul 4); în exemplul dat cele mai multe date provin din studii cvasiexperimentale ceea ce reduce, într-o oarecare măsură, nivelul calitativ al dovezilor privind eficienţa acestor forme de intervenţie;

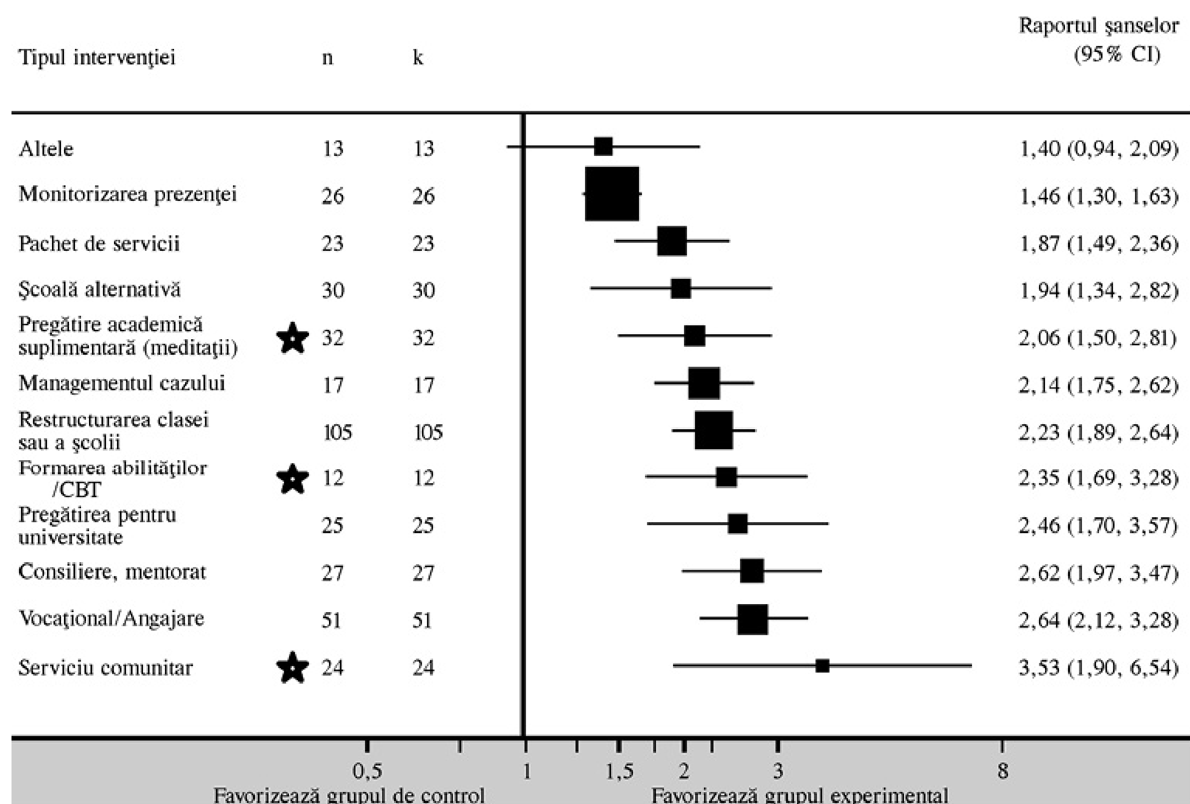


Figura 2.2. *Rezultatele metaanalizei efectuate de Wilson* et al. *(2011)*

Legendă: marcate cu steluţă sunt cele trei intervenţii de interes; n – număr de mărimi ale efectului; k – număr de studii independente; raportul şanselor (detalii legate de interpretarea metaanalizei în capitolul 7, secţiunile 7.2.1 şi 7.2.3);

(4) *luarea deciziei şi implementarea intervenţiei selectate*; în privinţa deciziei se au în vedere toate cele trei componente incluse într-o abordare de tip EBP (informaţii privind intervenţiile validate ştiinţific existente, privind caracteristicile şi valorile clientului şi din expertiza profesională a profesionistului); de exemplu, într-o asemenea decizie,

ţinând cont că orice formă de intervenţie necesită resurse, se poate lua în calcul nu numai eficienţa intervenţiei, ci şi un raport cost-eficienţă; de asemenea se poate analiza gradul de similaritate între contextul în care urmează să se implementeze intervenţia selectată şi cel al studiilor în care a fost validată intervenţia respectivă; astfel, dacă 90% dintre studiile care implementau programe educaţionale non-formale cu implicarea comunităţii vizau elevi proveniţi din familii de imigranţi de primă generaţie din comunităţi multiculturale, acestea s-ar putea să fie inadecvate pentru a fi aplicate în prevenţia abandonului şcolar din satele româneşti; nici varianta dezvoltării de abilităţi psihologice nu ar fi posibilă într-un asemenea context din lipsa unui personal calificat în acest mediu pentru aplicarea intervenţiilor psihologice; pe scurt, în judecarea situaţiei practice întâlnite se va ţine cont atât de specificul contextual, cât şi de dovezile ştiinţifice existente cu privire la eficienţa intervenţiei respective; în sfârşit, în privinţa implementării soluţiei selectate, acest lucru ar trebui să fie realizat în baza unui ghid (protocol) de aplicare, pentru a maximiza şansele de reproducere a efectelor pozitive din studiile de specialitate;

(5) monitorizarea şi evaluarea eficienţei intervenţiei; în urma aplicării intervenţiei este important să se urmărească dacă şi-a atins scopul pentru care a fost implementată, precum şi ce dificultăţi (implicit măsuri corective) au fost identificate ca fiind necesare în aplicarea intervenţiei.

Din nefericire, există o discrepanţă mare între acest algoritm prescriptiv de lucru, specific EBP, şi realitatea de zi cu zi. Deşi o ştiinţă matură ar trebui să se bazeze mai mult pe protocoale standard de intervenţie (vezi modelul EBP promovat în medicină) decât pe impresii şi

intuiţii personale, se întâmplă adesea ca măsurile de intervenţie luate de mulţi specialişti din domeniul psihologiei să fie bazate pe soluţii decise intuitiv, fără o analiză aprofundată a celor mai bune practici din domeniu. Acest lucru este facilitat şi de existenţa unor bariere în aplicarea modelului EBP, cele mai importante fiind sintetizate în tabelul 2.3.

Tabelul 2.3. *Bariere întâlnite în aplicarea modelului EBP în practica psihologică*

Barieră	*Descriere*
Timp limitat	Uneori li se solicită psihologilor să ofere ad-hoc recomandări practice. Într-o abordare EBP ar fi nevoie, de cele mai multe ori, de timp pentru a consulta literatura de specialitate şi a identifica intervenţii validate ştiinţific pentru problematica respectivă.
Acces limitat la resurse	În contextul în care se publică atât de mult, iar cea mai importantă bază de date din domeniul psihologiei (PsycInfo) este accesibilă doar în regim contra cost, practicienilor le este greu să realizeze o evaluare adecvată a literaturii de specialitate.
Abilităţi reduse de cercetare	Multor practicieni le lipsesc cunoştinţele necesare pentru a analiza critic literatura de specialitate, inclusiv la nivelul de a diferenţia între studii în funcţie de calitatea dovezilor ştiinţifice. Aceste abilităţi pot fi dobândite în special prin programe doctorale sau prin programe masterale serioase, însă nu toţi practicienii au un asemenea nivel de pregătire profesională.
Presiune scăzută de profesionalizare	Spre deosebire de alte domenii unde specializările sunt foarte precise, în domeniul serviciilor de psihologie sau conexe acesteia se întâmplă frecvent să existe consultanţi cu pregătire variată (ingineri, economişti etc.). În plus, deşi există o specializare inclusiv în domeniul psihologiei (de exemplu, psiholog clinician, psiholog educaţional etc.), există multe cazuri de practicieni cu atestat în multiple specialităţi, fiind greu de crezut faptul că aceştia pot performa în toate specialităţile pe care le deţin. De asemenea, erorile întâlnite în actul psihologic sunt rareori supuse unui demers de malpraxis, cum se întâmplă adesea în cazul serviciilor medicale, prin urmare există mai puţină presiune de tip punitiv cu privire la consecinţele deciziilor luate în practica psihologică.
Diversitatea conceptuală	În psihologie există o multitudine de direcţii teoretice cu un aparat conceptual propriu, fapt care conduce adesea la suprapuneri şi ambiguităţi terminologice. Această diversitate de opinii creează dificultăţi în acceptarea unui model comun,

pragmatic, cum este cel bazat pe EBP, mai ales din partea practicienilor adepţi ai unor perspective teoretice mai puţin fundamentate ştiinţific (empiric).

Dincolo de aceste bariere este esenţial pentru viitorul domeniului aplicativ al psihologiei să se adopte pe scală largă modelul EBP din motivele expuse în cele ce urmează.

Includerea serviciilor psihologice în cadrul a*sistenţei complexe interdisciplinare*. În cele mai multe cazuri, problemele oamenilor sunt abordate interdisciplinar. De exemplu, la ora actuală se tinde spre oferirea unor servicii integrate de sănătate (persoanele care au suferit un traumatism grav al membrului superior pot beneficia, pe lângă asistenţa medicală de specialitate, de servicii de kinetoterapie, precum şi de servicii psihologice pentru a facilita coping-ul adaptiv la noua situaţie şi pentru a identifica cele mai potrivite căi de reinserţie socioprofesională). În contextul în care profesiile din zona medicală au implementat deja modelul EBP, există o presiune suplimentară pentru practicienii din domeniul psihologiei de a-şi alinia soluţiile la aceeaşi paradigmă de lucru. Acest lucru nu se întâmplă doar în zona serviciilor de sănătate (unde modelul EBP este cel mai avansat), ci începe să fie întâlnit şi în alte domenii de activitate (educaţie – EBP în pedagogie; asistenţă socială – EBP în asistenţă socială; management – EBP în management etc.). Astfel, EBP oferă un cadru comun de raportare, fiind un model pragmatic şi ateoretic, bazat pe identificarea celor mai potrivite soluţii validate ştiinţific pentru o problemă dată.

Eficientizarea serviciilor psihologice. În mod curent (tradiţional), conform unor date din domeniul ştiinţelor medicale (*apud* Peterson, 2006), este necesar să treacă 17 ani de la publicarea rezultatelor cercetărilor cu privire la eficienţa unei forme de intervenţie până la includerea acesteia, într-o măsură relativ extinsă, în practica curentă.

Această distanţă este necesară pentru că adesea informaţiile provenite din articole de specialitate nu sunt citite direct de către practicieni. Aceştia din urmă preferă să îşi fundamenteze opiniile pe ceea ce cunosc deja din perioada facultăţii sau pe informaţii provenite din tratate fundamentale de specialitate, care sintetizează un domeniu anume. Însă de cele mai multe ori este nevoie de timp pentru ca un concept sau o formă de intervenţie anume să treacă din zona articolelor ştiinţifice, tributare unui limbaj tehnic şi statistic uneori greu de descifrat, în zona unor recomandări concrete transmise studenţilor în formarea lor academică sau specialiştilor în formarea lor profesională continuă. Prin abordarea EBP se solicită studiul aprofundat al literaturii de specialitate de ultimă oră (de exemplu, studii experimentale, metaanalize), iar acest aspect conduce la reducerea drastică a perioadei de timp necesară unei intervenţii pentru a intra în practica de zi cu zi. Acest lucru le permite practicienilor adepţi ai EBP să aibă un avantaj în faţa colegilor lor, printr-un plus de eficienţă la nivelul rezultatelor obţinute, ca urmare a accesului la cele mai recente/solide descoperiri ştiinţifice de interes pentru domeniul lor de expertiză şi a fundamentării practicii lor pe soluţii dovedite a fi eficiente[2].

Accesul la fondurile publice. Adaptând un slogan politic la modă, am putea spune că intervenţiile validate ştiinţific constituie o asigurare pentru contribuabil că taxele şi impozitele plătite către stat se regăsesc în accesul la servicii eficiente. De altfel, discursul autorităţilor în această direcţie este acela că resursele limitate existente ar trebui alocate către acele intervenţii pentru care există dovezi validate ştiinţific că sunt eficiente, că produc efectele scontate. Acesta este şi unul dintre motivele extinderii modelului EBP în domeniul sănătăţii, fiind facilitată de decizia autorităţilor din Marea Britanie, practică preluată apoi şi de alte state, de a aloca fonduri şi de a deconta servicii numai pentru acele intervenţii din

sfera sănătăţii pentru care există dovezi ştiinţifice solide cu privire la utilitatea lor (Spring, 2007). De aceea, şansa psihologilor de a-şi deconta serviciile prestate din fonduri publice constă în demersul de a dovedi că dispun de intervenţii validate ştiinţific pentru problemele apărute. Acest lucru este cu atât mai important cu cât aproape toate serviciile psihologice (educaţionale, clinice etc.), cu excepţia domeniului psihologiei muncii şi organizaţionale şi parţial a psihoterapiei (în unele state), sunt finanţate din fonduri publice. Însă nu numai autorităţile de stat sunt interesate să se implementeze cele mai eficiente forme de intervenţie, ci şi alţi finanţatori. De exemplu, companiile private de asigurări medicale din anumite ţări au stabilit că pot fi decontate servicii de psihologie dacă: (a) există dovezi empirice validate ştiinţific că intervenţia este una eficientă; (b) implementarea serviciului este una standard, pentru a permite asiguratorului să tragă la răspundere specialiştii care deviază nejustificat de la protocolul de implementare.

2.3. Informaţii şi resurse utile pentru practicienii care utilizează EBP

Dacă ar fi să ne referim la cele trei braţe ale modelului EBP (figura 2.1) vom constata adesea că practicienilor le este cel mai greu să intre în rolul de cercetător. Atât din aceste considerente, cât şi pentru faptul că lucrarea de faţă are drept obiectiv specific cercetarea psihologică, vom prezenta în cele ce urmează o serie de informaţii şi resurse utile pentru practicieni. Rolul acestora este de a facilita implementarea modelului EBP în practica profesională curentă.

Mai specific, vom urmări să oferim informaţii în legătură cu trei întrebări (aspecte) cheie din procesul EBP: (a) „ce surse de informare ar trebui accesate pentru a investiga care sunt intervenţiile validate ştiinţific pentru o anumită problemă psihologică?”; (b) „cât de valide sunt informaţiile respective/cât de mult ne putem încrede în ele?”; (c) „ce criterii ar trebui îndeplinite ca un serviciu psihologic să fie considerat ca fiind validat ştiinţific?”.

2.3.1. Surse de informare utile pentru practicieni

În urma exploziei de informaţii din era tehnologică, unui specialist în psihologie îi este tot mai greu să parcurgă toate studiile de specialitate dintr-un anumit domeniu. Pe lângă resursa de timp implicată, se poate pune şi problema costurilor, deoarece o parte însemnată din bazele de date relevante pentru psihologie sunt disponibile doar contra cost.

Luând în calcul motivele menţionate anterior, cel mai util pentru practicieni ar fi să caute dovezi privind nivelul de validare ştiinţifică în arhiva de studii *Cochrane Collaboration* (1993), utilă domeniilor din psihologie conexe ştiinţelor medicale (psihologie clinică, consiliere psihologică, psihoterapie), respectiv arhiva de studii *Campbell Collaboration* (2000), utilă pentru domenii ale psihologiei conexe ştiinţelor sociale (psihologia educaţională, psihologia juridică, psihologia comunitară, psihologia economică etc.). Recent, a fost gândită o asemenea bază de date şi în domeniul resurselor umane (psihologia muncii şi organizaţională), însă demersul se află abia la început, fiind greu de făcut predicţii cu privire la succesul acestei baze de date (*SIOP/SHRM Collaboration*, 2011).

Spre deosebire de celelalte baze de date întâlnite în domeniul cercetării cum ar fi PsycInfo, ISI Web of Science etc., cele amintite mai sus sunt axate pe sinteze de studii cu nivel ridicat de validitate ştiinţifică (metaanalize de studii experimentale sau cvasiexperimentale; sinteze narative) şi au o arie de adresabilitate extinsă (inclusiv practicienii cu un nivel mai scăzut de expertiză ştiinţifică), prin includerea de rezumate scrise într-un limbaj netehnic. De asemenea, accesul la consultarea sintezelor este liber (în regim gratuit)[3], oferind practicienilor o resursă serioasă spre consultare, în demersul de rezolvare a problemelor întâlnite în practica lor profesională. O selecţie a tematicilor abordate în cele două baze de date de tradiţie (Cochrane şi Campbell) este prezentată în caseta 2.2.

Caseta 2.2. *Exemple de tematici abordate în cele două baze de date*[4]

Exemple de sinteze disponibile în arhiva Cochrane Collaboration

1. Intervenţii eficiente în comunicarea către adolescenţi şi copii a diagnosticului de cancer acordat unui alt membru al familiei.
2. Eficienţa programelor de prevenire a fumatului derulate în şcoală.
3. Eficienţa terapiei de cuplu în combaterea depresiei.
4. Eficienţa terapiei prin realitate virtuală la pacienţii care au suferit un atac vascular cerebral.
5. Intervenţii cognitiv-comportamentale pentru prevenirea aderării copiilor şi adolescenţilor la *gangs* (grupuri de tineri cu potenţial infracţional ridicat).

Exemple de sinteze disponibile în arhiva Campbell Collaboration

1. Strategii eficiente de managementul clasei de elevi pentru reducerea comportamentelor agresive şi de indisciplină ale elevilor.
2. Eficienţa programelor de training adresate părinţilor cu copii diagnosticaţi cu ADHD.
3. Efectul programelor de tip *afterschool* asupra elevilor.
4. Eficienţa intervenţiilor de prevenire a accidentelor de la locul de muncă.
5. Eficienţa programelor de prevenire a abuzului sexual.

Numărul de sinteze disponibile este impresionant, îndeosebi a celor existente în arhiva *Cochrane Collaboration*. Cu toate că accesul la această arhivă din România este gratuit doar pentru rezumate, acestea sunt scrise într-un limbaj accesibil şi prezintă informaţii relevante care permit luarea unor decizii întemeiate pe studii validate ştiinţific. Asemenea informaţii sunt incomparabil mai valoroase (mai valide ştiinţific), decât dacă ne-am raporta la opiniile prezentate în cadrul unor forumuri de specialitate sau la analiza unui singur studiu empiric de specialitate[5].

Caseta 2.3. *Una din cele aproximativ 6.000 de sinteze disponibile în arhiva Cochrane*

Eficienţa usturoiului în tratarea răcelilor (Lissiman, Bhasale şi Cohen, 2009)
La nivel popular, usturoiul este perceput ca un tratament eficient în prevenirea şi tratarea răcelilor, probabil datorită studiilor de laborator care au arătat că usturoiul are proprietăţi antivirale şi antibacteriene. În analiză am căutat studii care să utilizeze usturoiul în prevenirea sau combaterea răcelii. Au fost identificate cinci studii, însă numai unul (N = 146 persoane) îndeplineşte criteriile pentru a fi inclus în analiză, fiind un studiu randomizat controlat.
Rezultate: cei care au consumat zilnic usturoi timp de trei luni de zile au avut un număr semnificativ statistic mai mic de răceli în comparaţie cu cei alocaţi în grupul placebo. În cazul participanţilor care au răcit, nu au existat diferenţe semnificative cu privire la durata episodului de răceală (4,63, respectiv 5,63 zile). Deşi rezultatul acestui studiu sprijină ideea rolului benefic al usturoiului în prevenirea apariţiei răcelilor, ar fi nevoie de mai multe studii randomizate pentru a replica rezultatele obţinute. Posibilele efecte secundare în acest studiu se referă la miros şi la unele iritaţii ale pielii. Studii suplimentare sunt necesare pentru a evidenţia clar efectele secundare în urma consumului de usturoi.

Desigur, chiar dacă în cele două baze de date cu tradiţie există un număr mare de sinteze este foarte posibil ca acestea să nu acopere toate posibilele probleme cu care se confruntă un practician. O alternativă utilă în acest caz, dar cu aplicabilitate restrânsă doar la domeniul psihoterapiei, este consultarea listei de intervenţii validate ştiinţific. Aceasta este actualizată periodic de un comitet din cadrul diviziei 12 a Asociaţiei

Psihologilor Americani. Lista este disponibilă spre consultare accesând linkul asociat referinţei bibliografice menţionate (Chembless *et al.*, 1996; 1998). În baza acestui demers, intervenţiile sunt grupate în două mari categorii: intervenţii validate ştiinţific, respectiv intervenţii cu probabilitate ridicată de a fi valide ştiinţific[6]. Totuşi, ar fi fost util să se introducă şi o a treia categorie, aceea a intervenţiilor despre care există dovezi că nu sunt valide ştiinţific. În lipsa acestei categorii nu se poate distinge dacă intervenţiile care nu sunt menţionate în cele două categorii sunt omise pentru că nu există studii de validare (nu există date) sau pentru că există studii, dar acestea un arătat că terapia respectivă este ineficientă. Un alt aspect important este faptul că împărţirea pe categorii se face în funcţie de problematica studiată, astfel încât este posibil ca aceeaşi formă de psihoterapie să fie eficientă pentru cazurile de bulimie, dar să nu fie inclusă în lista intervenţiilor validate ştiinţific sau a celor care au o probabilitate ridicată de a fi validate ştiinţific pentru intervenţii în cazuri de alcoolism.

O altă resursă avută la dispoziţie de către practicieni pentru a identifica intervenţii validate ştiinţific este apelul la baze de date internaţionale academice, precum PsycInfo, ERIC, PubMed etc., o parte dintre acestea fiind gratuite[7]. Dezavantajul apelării la aceste baze de date este faptul că, în cele mai multe cazuri, accesul gratuit este asigurat doar pentru studiul rezumatelor. Un alt neajuns este limbajul extrem de tehnic întâlnit, ceea ce presupune cunoştinţe metodologice şi statistice avansate (vezi capitolele 4-8).

În sfârşit, o altă sursă de informare utilă, mai ales pentru psihoterapeuţi şi psihologi clinicieni este apelul la ghiduri şi recomandări de derulare a activităţii profesionale pentru o tematică anume[8]. Din acest punct de vedere, recomandările NICE (2005) (acronim al organizaţiei

britanice *National Institute for Health and Clinical Excellence*) evidenţiază nivelul de încredere (validare ştiinţifică) în diferitele practici necesare în tratamentul unei probleme (de exemplu, recomandări pentru tratarea tulburărilor de comportament alimentar), dar şi informaţii cu privire la raportul cost-eficienţă sau cost-utilitate al unei intervenţii care s-a dovedit a fi eficientă din punct de vedere clinic.

2.3.2. Informaţii utile pentru evaluarea calităţii ştiinţifice a dovezilor existente

În concepţia publicului larg, dar din păcate şi a multor practicieni din domeniul psihologiei, cercetările conduc la adevăruri absolute, de necontestat. De aceea, frecvent în discursul argumentativ al multor persoane se face apel la rezultatele unor cercetări ce sprijină punctul lor de vedere. În realitate, lucrurile nu stau deloc aşa. Orice studiu, oricât de bine ar fi fost întocmit, are o serie de limite. De aceea este necesar să trecem prin filtrul gândirii critice rezultatele oricărui studiu. Mai mult, studiile diferă la nivelul validităţii concluziilor extrase din ele, cu alte cuvinte au niveluri diferite de validare ştiinţifică.

Deşi au fost dezvoltate mai multe variante ale piramidei încrederii în dovezile empirice rezultate din studii (cea mai cunoscută fiind dezvoltată de Sackett *et al.*, 2000), în cele ce urmează este prezentată o versiune proprie privind ierarhizarea dovezilor din perspectiva calităţii lor şi a nivelului de încredere avut în rezultatele studiului[9, 10]. Am ales această variantă pentru a integra informaţiile specifice legate de tipuri de cercetări, multe dintre ele fiind prezentate detaliat începând cu capitolul 4.

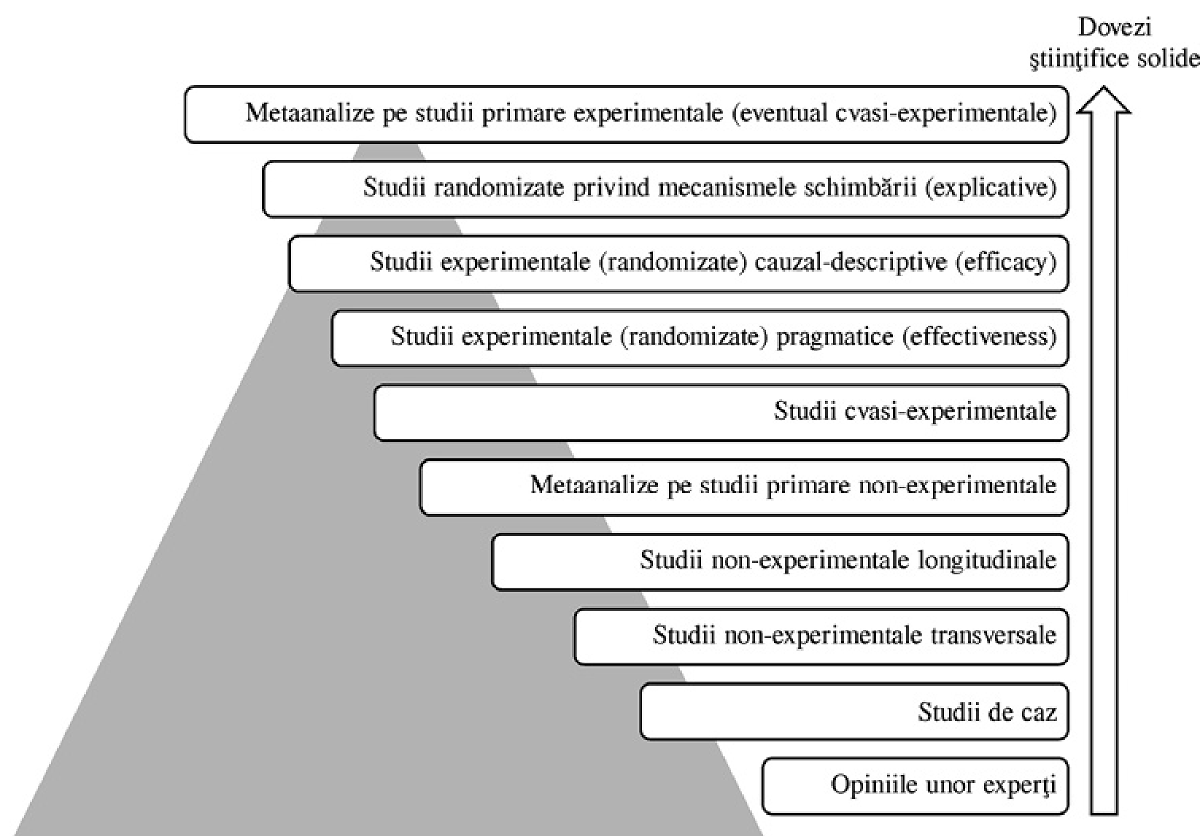


Figura 2.3. *Piramida încrederii privind eficienţa unei intervenţii psihologice în funcţie de calitatea metodologică a studiilor pe care se bazează*

Cum se interpretează un asemenea grafic? În primul rând, orice consumator al literaturii de specialitate ar putea să înţeleagă faptul că nivelul de încredere în concluziile unui studiu este mai ridicat atunci când există dovezi ştiinţifice solide. Dovezile sunt mai bune în cazul rezultatelor obţinute din studii aflate în partea superioară a piramidei (mai apropiate de vârful piramidei).

Să luăm exemplul ipotetic a unui specialist din domeniul psihologiei educaţionale care, în baza experienţei acumulate (observaţii şi reflecţii personale), şi-ar forma opinia că stima de sine influenţează nivelul epuizării emoţionale a profesorilor (*burnout*), în sensul că un nivel scăzut al stimei de sine la profesori conduce la intensificarea simptomatologiei asociate stării de epuizare emoţională a acestora. O asemenea opinie, dacă

nu este sprijinită de date provenite din studii de specialitate rămâne la un nivel scăzut de încredere (nivelul 1 – nivelul bazal din piramidă), lipsind dovezile care să o susţină. Chiar dacă ar exista unele dovezi ştiinţifice, cum ar fi cazul unui studiu corelaţional (nivelul 3 din piramidă – studii non-experimentale transversale), existenţa unei corelaţii negative semnificative statistic între stima de sine şi nivelul de *burnout* nu constituie o dovadă ştiinţifică suficientă pentru a afirma că stima de sine influenţează nivelul epuizării emoţionale, pentru că o asemenea afirmaţie specifică explicit existenţa unei relaţii de tip cauză-efect. Se ştie însă că relaţia de cauzalitate este numai una dintre variantele care poate conduce la o asociere semnificativă între cele două variabile. De exemplu, este posibil ca atât stima de sine scăzută, cât şi epuizarea emoţională, să fie amândouă consecinţe ale unui leadership autoritar din partea conducerii şcolii, orientat spre evidenţierea neregulilor corpului profesional (fără componenta feedbackului pozitiv). Acelaşi tip de relaţie poate fi pus pe seama altor variabile care ar putea juca un rol de tip cauză. Astfel, nivelul scăzut de autoeficacitate didactică ar însemna o încredere scăzută a profesorului în capacitatea sa de a face faţă la provocările profesiei de dascăl. Ar fi firesc, în aceste condiţii, ca acei profesori cu un nivel scăzut al autoeficacităţii didactice să manifeste niveluri mai scăzute ale stimei de sine şi un nivel mai ridicat de epuizare emoţională. O altă ipoteză plauzibilă pentru a explica rezultatul semnificativ statistic obţinut în studiul corelaţional ar fi chiar una opusă, cu privire la direcţia relaţiei cauzale. Astfel, nu ar fi de exclus ipoteza că epuizarea emoţională este sursa unui nivel mai scăzut al stimei de sine. De pildă, într-un context economic marcat de criză, în care există o lipsă acută de locuri de muncă vacante este posibil ca cineva care manifestă un nivel ridicat de epuizare emoţională să-şi dorească să caute un alt loc de muncă, dar să aibă

dificultăţi în găsirea unui alt loc mai bun. Această stare de blocaj s-ar putea răsfrânge în timp asupra nivelului stimei de sine. Pe scurt, toate aceste alternative explicative subliniază ideea că studiile corelaţionale pot valida doar afirmaţii ce indică ideea asocierii (există o corelaţie negativă între stima de sine şi nivelul epuizării emoţionale), însă nu pot conduce la afirmaţii valide cu privire la sursa (explicaţia) acestei asocieri etc.

Această situaţie se răsfrânge inclusiv la nivelul soluţiilor (intervenţiilor propuse). În lipsa unor dovezi ştiinţifice suficiente care să ateste relaţia de tip cauză-efect dintre stima de sine şi nivelul epuizării emoţionale este considerat inoportun să se implementeze un program de întărire a stimei de sine cu scopul de a reduce nivelul epuizării emoţionale.

Un pas suplimentar în discutarea relaţiilor de tip cauză-efect îl constituie studiile non-experimentale longitudinale (derulate pe o perioadă mai lungă de timp). În exemplul dat se poate institui drept criteriu de intrare măsurarea nivelului epuizării emoţionale la momentul iniţial (T0). Cei care obţin valori ridicate (ce indică prezenţa epuizării emoţionale) sunt excluşi din studiu. Astfel, la începutul studiului, toţii participanţii sunt încadraţi în categoria celor care nu sunt epuizaţi emoţional. Însă acest nivel al epuizării emoţionale este măsurat repetat, la diferite intervale de timp, şi se poate vedea astfel în ce măsură nivelul stimei de sine de la momentul anterior (de exemplu, T0) prezice dacă persoana este epuizată emoţional la următoarea testare (de exemplu, T1) sau mai ales în ce măsură modificări ale nivelului stimei de sine dintre momentele T0 şi T1 precedă modificări ale nivelului epuizării emoţionale între momentele T1 şi T2. Chiar dacă este vorba tot de un studiu non-experimental, designul longitudinal permite, între anumite limite importante, investigarea relaţiilor de tip cauză-efect, deoarece am putea

examina dacă modificările privind epuizarea emoţională (efectul) sunt precedate de modificări ale nivelului stimei de sine (cauza potenţială). Acest lucru este posibil deoarece una dintre condiţiile demonstrării unei relaţii de tip cauză-efect este aceea ca modificările la nivelul variabilei cauză să preceadă modificările la nivelul variabilei efect. O situaţie inversă, în care mai întâi ar apărea modificări la nivelul variabilei efect, ar fi o dovadă clară că relaţia cauză-efect propusă nu se susţine (nu este validată empiric).

Însă, demonstrarea antecedenţei temporale a variabilei cauză faţă de variabila-efect, deşi este o condiţie necesară, ea nu este şi una suficientă pentru a demonstra existenţa unei relaţii de tip cauză-efect. Revenind la exemplul dat, este posibil ca alte variabile străine să fie responsabile pentru variaţii ale nivelului epuizării emoţionale de la nivelul T1 la nivelul T2. De exemplu, pot exista şi alte variabile alternative (nivelul de autoeficacitate didactică, nivelul angajamentului organizaţional etc.) care să se modifice între momentele T0 şi T1. De asemenea, pot apărea unele evenimente importante între momentele testării care să fie responsabile pentru modificări la nivelul epuizării emoţionale. De exemplu, deşi modificările de la momentul T1 la momentul T2 sunt precedate de modificări ale stimei de sine de la momentul T0 la nivelul T1 este posibil să apară un eveniment deosebit între momentele T1 şi T2 (cum ar fi o lege de diminuare a salariului pentru personalul bugetar cu 25%) care să fie cauza reală a modificărilor observate cu privire la epuizarea emoţională. De aceea, studiile non-experimentale longitudinale, deşi constituie un pas suplimentar faţă de studiile non-experimentale transversale în privinţa dovezilor referitoare la o presupusă relaţie de tip cauză-efect, nu au acelaşi nivel de validitate privind detectarea unei relaţii de tip cauză-efect precum cel întâlnit în studiile experimentale, deoarece există un control

redus asupra variabilelor străine care ar putea afecta concluziile studiului. În schimb, dovezile obţinute din studiile experimentale sunt cele mai solide pentru a evidenţia o posibilă relaţie de tip cauză-efect deoarece, prin specificul lor, exercită cel mai bun control asupra variabilelor străine, care ar putea afecta concluziile unui studiu.

În capitolul 4 sunt detaliate diferenţele dintre diferitele tipuri de studii experimentale sau cvasiexperimentale, precum şi limitele derivate din specificul fiecăruia, iar în capitolul 5 sunt oferite detalii despre diferitele tipuri de studii non-experimentale. O scurtă descriere a fiecărui tip de studiu inclus în piramidă (figura 2.3) este prezentată în tabelul 2.4. Excepţie face cazul opiniilor situate la baza piramidei, acestea nefiind studii efective, ci impresii subiective, formate în timp în baza unui amalgam de experienţe personale şi de informaţii disparate provenite din studii.

Tabelul 2.4. *Descrierea succintă a principalelor tipuri de studii de interes pentru practicienii adepţi ai abordării EBP*

Denumire	Descriere succintă	Observaţii
Studii de caz	Analiza intensivă a unei unităţi (de exemplu, individ, grup, organizaţie) utilizate, fie cu scopul de a genera sau testa ipoteze (experimente pe un singur subiect), fie de a descrie (ilustra) impactul unei anumite teorii asupra unei situaţii particulare, contextuale (asupra unui individ).	Detalii în David (2006), Vîrgă (2007) şi în secţiunea 5.1
Studii non-experimentale transversale	Este cea mai întâlnită şi mai variată categorie de studii. Acestea pot fi împărţite în studii descriptive şi în studii analitice (corelaţionale, comparative, metodologice). Toate au ca element comun faptul că datele sunt colectate în aceeaşi unitate de timp şi că nu există o manipulare a variabilelor din partea cercetătorului. De aceea ele sunt cunoscute şi sub denumirea de studii observaţionale.	Detalii în capitolul 5, destinat studiilor non-experimentale

Studii non-experimentale longitudinale	Este o categorie aparte de studii observaţionale, cu deosebirea că datele sunt colectate în baza unor măsurători repetate pe acelaşi lot de persoane, întinse pe o perioadă mai îndelungată de timp. Fiind analizate aceleaşi persoane există un control mai bun asupra variabilelor străine decât în cazul studiilor transversale (sunt eliminate diferenţele interindividuale).	
Metaanaliză pe studii non-experimentale	Constă într-o analiză sistematică a studiilor anterioare de tip non-experimental. Principalul scop al metaanalizei este de a sintetiza rezultatele obţinute într-un domeniu, atât pentru a reflecta situaţia globală (în cazul în care studiile incluse prezintă rezultate contradictorii), cât şi pentru a surprinde acele variabile moderatoare care prezic variaţii ale efectelor observate. Deşi are avantajul cumulării studiilor, rezultatele obţinute rămân tributare limitelor studiilor non-experimentale incluse în analiză.	Detalii în capitolul 7, destinat studiilor metaanalitice
Studii cvasiexperimentale	Acest gen de studii implică manipularea variabilei independente (VI) pentru a măsura impactul acesteia asupra variabilei dependente, însă grupurile comparate nu sunt formate prin alocarea randomizată a participanţilor, fapt care afectează calitatea controlului exercitat asupra variabilelor străine.	Detalii în capitolul 4, destinat studiilor experimentale şi celor cvasiexperimentale
Studii experimentale pragmatice	Acestea îndeplinesc ambele cerinţe necesare pentru un studiu experimental: manipularea VI şi randomizarea participanţilor în grupele experimentale. Totuşi, unele variabile pot influenţa rezultatele studiului, nefiind exercitat în mod deliberat un control strict asupra altor variabile pentru a vedea dacă efectul anticipat se manifestă în condiţiile naturale normale (de exemplu, criterii mai permisive de includere a participanţilor în studiu, importanţă scăzută acordată standardizării intervenţiei etc.).	
Studii experimentale cauzal-descriptive	Acestea solicită un control strict al tuturor variabilelor care ar putea afecta concluziile studiului, fiind un design mai strict comparativ cu cel întâlnit în studiile pragmatice. Pe de altă parte, doar pe baza acestui tip de design se poate dovedi clar existenţa unei relaţii de tip cauză-efect.	

Studii experimentale asupra mecanismelor schimbării	Deşi este important să se demonstreze clar relaţia de tip cauză-efect (de exemplu, că terapia familial-sistemică este eficientă pentru a îmbunătăţi satisfacţia maritală), ar fi poate mai important să identificăm ce anume din cadrul intervenţiei a condus la creşterea satisfacţiei maritale, mai exact care au fost cauzele (mecanismele schimbării). Dacă se cunoaşte acest lucru pot fi optimizate ulterior intervenţiile pentru a maximiza efectul dorit.	
Metaanaliză pe studii experimentale	Constă în analiza sistematică a studiilor anterioare de tip experimental (eventual şi a celor de tip cvasiexperimental). Rezultatele acestui tip de studii constituie cele mai solide dovezi ştiinţifice pentru că beneficiază atât de avantajele criteriului cantitativ (analiza mai multor studii), cât şi de cele ale criteriului calitativ (fiind incluse doar studii experimentale, acestea având cel mai scăzut risc de a conduce la concluzii distorsionate).	Detalii în capitolul 7, destinat studiilor metaanalitice

Un alt mod de a ne raporta la piramida exprimată în figura 2.3 este că, din punct de vedere cantitativ, cele mai multe dovezi provin din partea de jos a piramidei. Din păcate, opiniile sunt cele mai frecvent întâlnite în practica de zi cu zi. De regulă există mult mai multe opinii şi studii de tip non-experimental referitoare la o problemă decât există studii de tip experimental pe aceeaşi problematică. De aceea, cu cât se urcă spre vârful piramidei există tot mai puţine studii disponibile. Se pot detecta mai multe motive pentru această stare de fapt, dintre care două sunt cel mai frecvent întâlnite. Primul motiv ar fi acela că studiile de tip experimental sunt studii care se fac pe probleme „mature”, adică asupra unor subiecte despre care există deja suficientă informaţie ştiinţifică provenită din studii non-experimentale. Deoarece studiile experimentale necesită multe resurse logistice, nu ar fi normal, de exemplu, să se deruleze un studiu experimental care să urmărească efectul terapiei raţional emotive (REBT)

asupra anxietăţii sociale, înainte de a exista dovezi că există o corelaţie între nivelul anxietăţii sociale şi nivelul cogniţiilor iraţionale (acestea fiind ţinta directă a modificărilor prin intermediul REBT). Cel de-al doilea motiv este de ordin etic. Pentru a derula studii experimentale sau cvasiexperimentale este necesar să se realizeze o manipulare a variabilei independente de către cercetător, însă acest lucru nu este întotdeauna posibil din motive etice. Astfel, nu ar fi acceptabil să facem un experiment cu privire la impactul abuzului emoţional exercitat asupra partenerului de cuplu asupra nivelului de depresie. Ceea ce putem face în acest caz este să ne bazăm pe studii de tip non-experimental pentru că nu ar fi etic să împărţim randomizat participanţii în două grupe, una experimentală şi una de control, iar celor din grupa experimentală să le solicităm (să-i încurajăm) să manifeste comportamente abuzive din punct de vedere emoţional.

Pentru a complica şi mai mult lucrurile legate de nivelul de încredere în rezultatele diferitelor studii, se cuvine să precizăm că pot exista diferenţe calitative inclusiv între studiile aflate în aceeaşi categorie (pe aceeaşi treaptă) din piramidă. De exemplu, în cazul studiilor randomizate de tip descriptiv-cauzal (*efficacy studies*) pot exista diferenţe, de exemplu, în funcţie de caracteristicile grupului de control. Astfel, în general se acordă o mai mare încredere în rezultatele studiilor dacă grupul experimental este comparat cu un grup de control care primeşte un tratament de tip placebo[11], decât dacă grupul experimental este comparat cu un grup de control care momentan nu primeşte vreun tratament, fiind pe lista de aşteptare (*waiting list*). Acest lucru se întâmplă deoarece în primul caz putem distinge clar cât din efectul observat se datorează intervenţiei şi cât este doar efect placebo[12], pe când în cel de-al doilea caz nu este foarte clar dacă eventualele diferenţe observate sunt cauzate de

intervenţia în sine sau de efectul placebo (vezi de asemenea efectul Hawthorne).

De aceea, pentru a surprinde şi diferenţe calitative dintre studii, altele decât cele obţinute prin raportarea la piramida calităţii dovezilor ştiinţifice, au fost instituite unele grile de evaluare a calităţii dovezilor ştiinţifice care să includă şi alte criterii de analiză. Una dintre cele mai răspândite grile de evaluare de acest gen este sistemul GRADE[13] (Atkins *et al.*, 2004; Balshem *et al.*, 2011; Guyatt *et al.*, 2011) care include, pe lângă (a) designul studiului ce se regăseşte şi în piramida calităţii dovezilor ştiinţifice, informaţii despre: (b) calitatea studiilor; (c) consistenţa rezultatelor; (d) similaritatea studiilor cu situaţia practică de rezolvat. În baza datelor referitoare la design, la calitatea studiilor, la robusteţea şi omogenitatea efectelor şi la gradul de aplicabilitate a acestora la situaţia practică întâlnită, sistemul GRADE clasifică intervenţiile în patru categorii diferite, din perspectiva gradului de încredere că acestea vor funcţiona pentru situaţia practică întâlnită (caseta 2.4).

Caseta 2.4. *Principalele atribute ale sistemului GRADE (Balshem* et al.*, 2011)*

Niveluri de încredere în reuşita intervenţiei (punctaj între paranteze):

- ridicat (4) – probabilitate scăzută ca studii viitoare să conducă la modificarea încrederii în eficienţa intervenţiei;
- moderat (3) – studii viitoare pot avea un impact relevant pentru concluziile referitoare la eficienţa intervenţiei;
- scăzut (2) – probabilitate ridicată ca studiile viitoare să aibă impact important asupra concluziilor, fiind posibil să asistăm la situaţia inversării direcţiei efectului;
- foarte scăzut (1) – orice direcţie a efectului este posibilă.

Scoruri de pornire iniţială:

- dovezi obţinute în baza unor studii experimentale – 4;
- dovezi obţinute în baza unor studii non-experimentale (observaţionale) sau cvasiexperimentale – 2;

- dovezi bazate pe opinii – 1.

Aceste punctaje sunt ajustate după cum urmează:

Scăderi de punctaj

- –1 sau –2 puncte pentru calitatea scăzută a studiilor sub standardul implicat de designul respectiv;
- –1 pentru număr mic de studii sau pentru cazul în care se observă efecte eterogene (variaţii mari);
- –1 sau –2 puncte penalizare pentru studiile care au puţine similitudini cu grupul-ţintă vizat în practică;
- –1 pentru situaţii de prezentare trunchiată a rezultatelor (*reporting bias*).

Creşteri de punctaj

- +1 dacă variabilele străine necontrolate ar fi putut afecta datele doar în direcţia reducerii efectului;
- +1 dacă este manipulată VI (studii cvasiexperimentale);
- +1 sau +2 dacă există asocieri puternice între variabile (risc relativ dublu, Cohen d cu aproximaţie de 0,40), respectiv foarte puternice (risc relativ de 5 ori mai mare/mai mic, Cohen d aproximativ 0,90).

Avantajul utilizării GRADE este acela că, în anumite cazuri, informaţii poziţionate inferior pe scala calităţii dovezilor din perspectiva designului pot fi considerate mai utile decât cele situate pe poziţii superioare. Un astfel de exemplu, ar fi cazul comparării a două forme de psihoterapie A şi B pentru combaterea depresiei majore la vârstnici. Intervenţia A este considerată mai potrivită de utilizat (dacă ea se bazează pe eficienţa observată în două studii de tip cvasiexperimental, ambele efectuate pe o populaţie similară ca vârstă şi din punct de vedere cultural şi în condiţiile în care se observă efecte omogene şi robuste), decât terapia B (pentru care există dovezi privind eficienţa acesteia într-un studiu experimental pragmatic, dar efectuat pe o populaţie diferită, de exemplu, pe tineri delincvenţi).[14]

Utilizarea sistemului GRADE, deşi nu specifică explicit ce înseamnă o intervenţie validată ştiinţific, constituie un reper util pentru practicieni,

fiind indicat să se apeleze doar la intervenţii catalogate ca având un nivel de încredere ridicat sau unul moderat.

2.3.3. Dovezi ştiinţifice minimale pentru a considera o intervenţie drept validată ştiinţific

În urma căutării informaţiilor relevante din literatura de specialitate şi a analizei critice privind calitatea şi relevanţa lor pentru problema investigată, se cuvine să precizăm nişte standarde (*benchmarks*) pentru a considera o anumită intervenţie din domeniul psihologiei, ca fiind validată ştiinţific.

Asociaţia Psihologilor Americani (APA), prin divizia de Psihologie Clinică şi Psihoterapie (vezi Chambless *et al.*, 1996, 1998 sau Chambless şi Hollon, 1998) propune asemenea repere pentru segmentul de intervenţii terapeutice. Conform acestei propuneri, există două categorii acceptabile de intervenţii terapeutice validate ştiinţific: intervenţii a căror eficienţă este clar demonstrată, respectiv intervenţii probabil eficiente. Aceste repere constituie o poziţie oficială pentru domeniul psihoterapiei, fiind la baza alcătuirii listei de intervenţii validate ştiinţific realizată şi actualizată periodic de către specialiştii APA.

Caseta 2.5. *Criterii minimale pentru intervenţiile validate ştiinţific*

Intervenţii eficiente (*well-established treatment):*

1. cel puţin două studii experimentale de tip cauzal-descriptiv (*efficacy studies*) în care să fie demonstrată eficienţa intervenţiei în unul sau mai multe dintre următoarele moduri:
   a) intervenţia este superioară unei intervenţii psihologice de tip placebo sau unei alte intervenţii psihoterapeutice sau din zona farmacoterapiei;
   b) intervenţia este echivalentă unei alte intervenţii psihoterapeutice a cărei validitate a fost demonstrată anterior, în cadrul unor studii care au o putere

statistică adecvată (de exemplu, cel puţin 30 de persoane per grup experimental);

sau

2. un număr mare (n ≥ 9) de experimente pe un singur participant (*single-case experiments*) în care s-a constatat că intervenţia este una eficientă, utilizând un design experimental adecvat şi respectând criteriul 1.a.

Condiţii suplimentare solicitate atât pentru cazul 1, cât şi pentru cazul 2:

3. intervenţiile să fie implementate în baza unui manual (să fie standardizate);
4. criteriile de includere (excludere) a participanţilor în studiu să fie explicitate, astfel încât să fie clare caracteristicile participanţilor;
5. eficienţa intervenţiei să fie demonstrată de cel puţin doi cercetători diferiţi (echipe de cercetare diferite).

Intervenţii probabil eficiente:

1. două studii experimentale în care se demonstrează că intervenţia este eficientă comparativ cu un grup de control tip listă de aşteptare (care nu beneficiază momentan de niciun fel de intervenţie);

sau

2. un studiu experimental care îndeplineşte criteriile 1, 3 şi 4 menţionate în categoria intervenţiilor eficiente;

sau

3. un număr mai redus de experimente pe un singur subiect (n ≥ 3) derulate cu respectarea criteriilor 2, 3 şi 4 menţionate în categoria intervenţiilor eficiente.

Notă: Evaluarea se face separat pentru fiecare tip de problemă în parte. Spre exemplu, unele intervenţii pot fi considerate eficiente pentru tratarea depresiei majore, pot fi probabil eficiente pentru tratarea anxietăţii sociale şi pot fi chiar ineficiente (sau pot lipsi dovezi suficiente legate de eficienţa lor) pentru tratarea atacurilor de panică.

De-a lungul timpului au fost dezvoltate mai multe asemenea criterii minimale (vezi lista lor în Chambless şi Ollendick, 2001 sau în Kratochwill şi Stoiber, 2002). La acestea aş adăuga contribuţia recentă adusă de profesorul clujean Daniel David împreună cu cercetătorul american Guy Montgomery. Aceştia propun o altă grilă pentru încadrarea intervenţiilor în categoria celor validate ştiinţific, punând accent nu numai

pe eficienţa unei intervenţii anume, ci şi pe existenţa unor dovezi privind mecanismele cauzale responsabile pentru această schimbare (David şi Montgomery, 2011). Cei doi autori au pornit în demersul lor de la: (a) problema existenţei multiplelor cadre de referinţă existente pentru a încadra o intervenţie ca fiind validată ştiinţific, ceea ce conduce la situaţia neplăcută ca anumite intervenţii psihoterapeutice să fie considerate validate ştiinţific după anumite criterii, dar nu şi după altele şi (b) problema accentului excesiv pus pe eficienţa intervenţiei cu neglijarea cadrului teoretic. În consecinţă, cei doi autori propun să se diferenţieze între intervenţiile psihoterapeutice nu numai în funcţie de eficienţa lor, ci şi în funcţie de fundamentarea ştiinţifică a teoriei de la baza lor.

Astfel, David şi Montgomery (2011) propun un sistem de evaluare format din nouă clase (categorii) de intervenţii psihoterapeutice rezultate din intersectarea a două criterii majore, fiecare cu câte trei niveluri de raportare. Cele două criterii sunt: eficienţa terapeutică (1 – efecte benefice; 2 – rezultate neclare/contradictorii/nedovedite; 3 – efecte absente/adverse); respectiv fundamentarea teoretică a intervenţiei (mecanismele cauzale implicate) (1 – dovezi clare; 2 – dovezi inconcludente/contradictorii/absenţa testării; 3 – dovezi care contrazic clar modelul teoretic/teorii pseudoştiinţifice, ce nu pot fi infirmate).

Cele nouă categorii sunt prezentate în figura 2.4. Dintre acestea, cinci categorii (clase) sunt considerate ca fiind invalide din punct de vedere ştiinţific pentru că, fie există dovezi privind lipsa lor de eficienţă terapeutică, fie se bazează pe teorii invalidate, fie au ambele neajunsuri menţionate (cum ar fi cazul intervenţiilor incluse în clasa 9). Se consideră că există dovezi despre o anumită tehnică dacă există rezultate similare obţinute din cel puţin două studii experimentale efectuate de cercetători sau echipe de cercetare diferite.

Desigur, ideal pentru orice intervenţie terapeutică ar fi să ajungă să fie încadrată în prima clasă, aceea a intervenţiilor pentru care există dovezi, atât cu privire la eficienţa lor, cât şi cu privire la cauzele (teoria) pentru care acestea sunt eficiente, prin identificarea mecanismelor schimbării. Pentru ca eficienţa să fie demonstrată este nevoie ca dovezile să provină din cel puţin două studii experimentale, efectuate de cercetători sau echipe de cercetare diferite. De remarcat, de asemenea, că orice intervenţie nouă (cu excepţia cazului când este fundamentată pe teorii invalidate ştiinţific sau pe modele pseudoştiinţifice), este încadrată pentru început în clasa 4, destinată intervenţiilor pentru care încă nu există suficiente dovezi privind eficienţa lor şi mecanismele cauzale implicate, dar care se bazează pe modele teoretice generale pentru care există suport ştiinţific.

	Mecanism cauzal susţinut	Informaţii insuficiente	Mecanism cauzal respins/imposibil de testat
Eficienţă demonstrată	Clasa 1	Clasa 2	Clasa 5
Rezultate contradictorii, eterogene/lipsă informaţii	Clasa 3	Clasa 4	Clasa 7
Eficienţă absentă/efecte adverse prezente	Clasa 6	Clasa 8	Clasa 9

Figura 2.4. *Clasificarea intervenţiilor după David şi Montgomery (2011), adaptare*

Deşi cele mai multe demersuri de evaluare a intervenţiilor aparţin domeniul psihologiei clinice şi psihoterapiei, trebuie menţionat că demersuri similare au fost efectuate şi în alte domenii aplicative ale psihologiei. De exemplu, Diviza 16 a APA a propus un sistem de lucru pentru stabilirea eficienţei intervenţiilor din domeniul psihologiei şcolare (Kratochwill şi Stoiber, 2002). Acesta este probabil cel mai comprehensiv sistem de evaluare (de exemplu, manualul de codare are nu mai puţin de

134 de pagini – pentru detalii, vezi Kratochwill, 2012), însă tocmai din cauza acestei complexităţi sporite este puţin probabil să fie adoptat în practica profesională. În schimb, sistemul este foarte util pentru cercetătorii interesaţi să stabilească nivelul de validare ştiinţifică pentru diferitele intervenţii din domeniul educaţional şi pentru practicienii interesaţi de a avea o imagine de ansamblu (în scop educaţional) cu privire la detaliile implicate într-o abordare de tip practician-cercetător (*scientist-practitioner*).

Acest model complex cuprinde, pe lângă caracteristicile precizate pentru aria psihologiei clinice (accentul pus pe evaluarea eficienţei unei intervenţii din perspectiva validităţii interne a studiilor), atât criterii ce ţin de validitatea externă a rezultatelor (similaritate culturală, de vârstă etc.), cât şi aspecte ce vizează validitatea de construct şi validitatea statistică a studiilor (detalii despre tipurile de validitate sunt prezentate în capitolul următor). O altă caracteristică specifică este includerea în procesul de evaluare a intervenţiilor în sprijinul cărora datele provin din studii calitative.

Ca modalitate de punctare, fiecare aspect este evaluat utilizând un sistem de codare în patru puncte (3 – dovezi puternice; 2 – dovezi promiţătoare; 1 – dovezi limitate [slabe]; 0 – absenţa dovezilor). De exemplu, în evaluarea calităţii studiilor cu referire la grupele de control utilizate se acordă: 3 puncte pentru grupuri de control active formate prin randomizare (intervenţii alternative, placebo etc.); 2 puncte pentru grupuri de control pasive formate prin randomizare (listă de aşteptare); 1 punct pentru grupuri de control formate fără randomizare (studii cvasiexperimentale), dar în care există demonstrată echivalenţa cu grupul experimental la nivel de pretest pentru variabilele de interes sau măcar există niveluri similare procentuale cu referire la retragerea din

experiment şi 0 puncte pentru situaţia absenţei unui grup de control sau pentru cazul unor grupuri de control pentru care nu s-a făcut niciun efort de asigurare a echivalenţei lor cu grupul experimental.

La ora actuală nu am cunoştinţă de existenţa vreunui demers de clasificare a intervenţiilor în zona psihologiei muncii şi organizaţională, însă principiile menţionate pentru cazurile psihoterapiei sau psihologiei şcolare pot fi extrapolate de către practicienii interesaţi şi pentru analiza intervenţiilor din domeniul organizaţional[15].

În schimb au fost formulate sugestii de clasificare a procesului de evaluare psihologică (toate celelalte clasificări menţionate referindu-se la partea de intervenţie psihologică) (Cohen *et al.*, 2008). Criteriile de bază în zona evaluării psihologice, aşa cum era de aşteptat, vizează în principal caracteristicile psihometrice ale metodei de evaluare (cel mai adesea un test psihologic).

Caseta 2.5. *Criterii utilizate pentru evaluarea instrumentelor psihologice din perspectiva EBP*

Metode solide (*well-established assessment)*

1. Instrumentul este prezentat în cel puţin două articole de specialitate cu *peer-review*, de către doi cercetători sau două echipe de cercetare diferite.
2. Sunt prezentate detaliat informaţii psihometrice care să ateste că metoda de evaluare are o bună validitate şi fidelitate în cel puţin un articol cu *peer-review*.
3. La cerere sunt oferite informaţii suplimentare (există un manual) care să prezinte alte informaţii relevante, încât să permită o evaluare critică a rezultatelor şi implementarea standard a metodei.

Metode probabil eficiente

1. Instrumentul este prezentat în cel puţin două articole de specialitate cu *peer-review*, de către acelaşi autor sau aceeaşi echipă de cercetare.
2. Sunt prezentate detaliat informaţii psihometrice care să ateste că metoda de evaluare are o validitate şi o fidelitate acceptabilă în cel puţin un articol cu *peer-review*.

3. La cerere sunt oferite informaţii suplimentare (există un manual) care să prezinte alte informaţii relevante, încât să permită o evaluare critică a rezultatelor şi implementarea standard a metodei.

Metode promiţătoare

1. Instrumentul este prezentat într-un singur articol de specialitate cu *peer-review*.
2. Caracteristicile psihometrice associate metodei sunt bune sau cel puţin adecvate.
3. La cerere sunt oferite informaţii suplimentare (există un manual) care să prezinte alte informaţii relevante, încât să permită o evaluare critică a rezultatelor şi implementarea standard a metodei.

O clasificare oarecum similară a fost adoptată de Colegiul Psihologilor din România (Liţă, Stan şi Sava, 2008), conform căreia se acordă: (a) aviz pe perioadă nedeterminată (metode solide) acordat probelor care dovedesc existenţa unui manual, dovezi numeroase şi robuste privind caracteristicile psihometrice (validitate şi fidelitate), precum şi dovada adaptării pe populaţie românească, inclusiv etaloane, dacă este cazul; (b) sau aviz pe perioadă determinată (metode probabil eficiente/promiţătoare) acordat probelor care dovedesc existenţa unui manual, prezintă dovezi suficiente privind validitatea testului şi acceptabile cu privire la fidelitate şi etalonare şi dovezi de adaptare pe populaţie românească, dacă este cazul. Probele care nu îndeplinesc aceste criterii minimale sunt respinse de la avizare. O listă a probelor psihologice avizate de către Comisia de Metodologie a Colegiului Psihologilor din România (CPR) este disponibilă la www.copsi.ro.

Din păcate nu au fost demarate asemenea demersuri de către alte comisii aplicative ale CPR pentru a identifica intervenţiile validate ştiinţific din diferite domenii de activitate, deşi acest lucru ar fi putut fi făcut, cel puţin în domeniul psihologiei clinice, a consilierii psihologice şi a psihoterapiei unde există multă literatură specifică.

2.3.4. Sugestii suplimentare privind luarea deciziei în practica EBP

La finalul acestui proces laborios pe care ar trebui să-l parcurgă un practician adept al practicii EBP apare problema selectării celei mai bune soluţii de implementat în practică.

Să ne reamintim că în EBP trebuie să se ţină seama, pe lângă identificarea unor intervenţii validate ştiinţific, şi de caracteristicile clientului, precum şi de nivelul de expertiză (existenţa competenţei necesare) a psihologului în a implementa soluţia cea mai bună. La aceasta se adaugă resursele existente, fiind preferate soluţiile care dovedesc că sunt eficiente şi fezabile de implementat, dar şi care posedă un foarte bun raport cost-eficienţă.

Deşi decizia în sine ţine de o abordare contextuală, ideografică, ar fi util ca orice practician să răspundă la o serie de întrebări (vezi tabelul 2.5, o adaptare după modelul sugerat de Kratochwill şi Shernoff, 2004), pentru a alege cea mai potrivită intervenţie pentru contextul respectiv.

Tabelul 2.5. *Zece întrebări-cheie (filtru) în selectarea intervenţiilor psihologice*

Dimensiuni vizate	Soluţia 1	Soluţia 2	Soluţia 3
1. Clientul are caracteristici similare cu cele întâlnite în soluţiile identificate ca fiind validate ştiinţific? *Caracteristici demografice* – vârstă, sex, nivel de educaţie, spaţiu cultural. *Solicitantul intervenţiei* – clientul, părintele, angajatorul etc.			
2. Cât de complexă este intervenţia (de reproductibilă)? *Specifică* – replicabilă uşor; *complexă* – reuşita ei depinde de mulţi factori			
3. Există factori contextuali (organizaţionali) care pot interfera cu privire la reuşita intervenţiei?			

Climat instituţional şi politici organizaţionale, sprijinul conducerii, percepţia beneficiarilor faţă de soluţia propusă
4. Există expertiza necesară pentru implementarea soluţiei? *Certificări necesare, experienţă anterioară*
5. Există acces la un manual sau măcar la un ghid de implementare a intervenţiei?
6. Poate fi încadrată intervenţia în categoria celor validate ştiinţific?
7. Există dovezi solide privind eficienţa intervenţiei propuse? Efecte omogene şi robuste (semnificative din punct de vedere clinic, practic)
8. Intervenţia constituie un factor de risc pentru apariţia unor efecte secundare adverse (pot fi afectate alte aspecte importante pentru client)?
9. Există indicatori obiectivi (operaţionalizabili) care să monitorizeze eficienţa intervenţiei, înainte de finalizarea ei?
10. Intervenţia prezintă un bun raport cost-eficienţă?

De asemenea, în contextul decizional implicat, practicianul ar trebui să fie atent să nu se lase influenţat de distorsiunile cognitive inerente situaţiilor de decizie şi de evaluare (monitorizare) a procesului de implementare. Printre cele mai frecvente se regăsesc: euristica de accesibilitate (indiferent de natura problemei alegem soluţia la îndemână, care a fost utilizată şi cu alte ocazii sau cea în care avem cea mai mare experienţă – „de ce nu încerci X, am obţinut multe rezultate bune”); euristica exemplului-clişeu (conturarea unui profil stereotipal – „vă cunosc eu pe voi «ăştia», copiii străzii, la toţi funcţionează metoda X”); biasul confirmării (amintirea situaţiilor care se pliază pe teoria implicită construită despre problema clientului, cu scop asigurator că lucrurile sunt pe o direcţie bună); apelul la explicaţii *post hoc* (justificări aduse unor situaţii apărute – „metoda funcţionează mai puţin eficient acum decât în alte ocazii pentru că au apărut unele circumstanţe agravante”).

În sfârşit, acolo unde este posibil ar trebui ca în implementarea intervenţiei selectate să se ţină cont de recomandările existente în ghidurile de intervenţie (în cazul în care sunt elaborate asemenea documente). Aceasta este dovada finală care distinge la nivel aplicativ între diferitele domenii de activitate. O ştiinţă matură ar trebui să se bazeze pe protocoale specifice de intervenţie şi mai puţin pe impresii personale şi opinii despre cum ar trebui să fie implementată o anumită intervenţie. De altfel, în medicină acest aspect face de multe ori diferenţa dintre praxis şi malpraxis şi este posibil ca, odată cu dezvoltarea mai multor ghiduri şi protocoale de intervenţie, să se ajungă la aceeaşi stare de fapt şi în psihologie.

Concluzii

Conceptul practicii validate ştiinţific (EBP) cunoaşte în ultimii 10 ani o dezvoltare deosebită, fiind introdus treptat în domenii aplicative variate (management, ştiinţele educaţiei, politici publice, psihologie, ştiinţe medicale). Ideea de bază este de a identifica şi implementa în practica de zi cu zi acele intervenţii care s-au dovedit a fi validate ştiinţific. Acceptarea acestora şi încrederea pe care o avem în eficienţa lor este dependentă de calitatea dovezilor ştiinţifice existente, în baza studiilor empirice realizate. Din acest punct de vedere, standardul spre care se tinde este acela de a obţine dovezi din studii randomizate (experimentale) şi, îndeosebi, din metaanalize care sintetizează rezultatele acestor studii experimentale.

EBP înseamnă mai mult decât intervenţii validate ştiinţific. Implică, de asemenea, a ţine cont de expertiza clinică a celui care implementează

soluţia şi de caracteristicile specifice ale clientului căruia i se propune intervenţia respectivă. Astfel, pe lângă dimensiunea ce vizează eficienţa intervenţiei, la fel de relevantă este şi dimensiunea ce vizează utilitatea aplicativă (în sfera sănătăţii se foloseşte denumirea de utilitate clinică). Aceasta se referă la discutarea fezabilităţii, aplicabilităţii şi a utilităţii (mărimea efectului) unei intervenţii validate ştiinţific, vizând astfel aspecte ce ţin de validitatea externă.

Adoptarea unui asemenea model profesionist din partea practicienilor, în ciuda unor inconvenienţe (necesită resurse de timp şi un efort mai mare), va conduce la multiple beneficii pentru profesia de psiholog prin: (a) o mai mare responsabilizare în legătură cu actul profesional prin respectarea unor protocoale de intervenţie; (b) o creştere a calităţii serviciilor psihologice prin obţinerea de efecte pozitive mai intense în urma implementării celor mai potrivite tehnici de intervenţie şi prin eliminarea, în timp, a practicilor demodate care s-au dovedit a fi ineficiente; (c) dezvoltarea pregătirii profesionale a psihologilor prin apariţia de materiale educaţionale şi comerciale diverse (manuale de intervenţie, casete video etc.) şi dezvoltarea psihologiei ca ştiinţă, prin creşterea nivelului de alfabetism ştiinţific printre practicieni, ca o măsură suplimentară menită să reducă discursul de tip pseudoştiinţific din domeniul psihologiei.

Lecturi suplimentare

Rob B. Briner şi Danise M. Rousseau (2011). Evidence-based I-O psychology; Not there yet. *Industrial and Organizational Psychology*, 4, pp. 3-22.

Articolul prezintă starea de fapt cu privire la situaţia practicii validate ştiinţific în domeniul psihologiei muncii şi organizaţională.

Bonnie Spring (2007). Evidence-based practice in clinical psychology: What it is, why it matters, what you need to know. *Journal of Clinical Psychology*, 63, pp. 611-631.

Articolul are marea calitate de a prezenta într-o manieră foarte structurată aspectele-cheie ale conceptului de EBP (istoric, definire, resurse) cu referire la domeniul psihologiei clinice şi al psihoterapiei.

Cochrane Collaboration – www.cochrane.org

Bază de date gratuită, al cărei scop este de a colecta şi evalua toate studiile randomizate din zona sănătăţii cu scopul de a-i ajuta pe practicieni şi pe cei care dezvoltă politici în domeniul sănătăţii să ia deciziile cele mai bune. La ora actuală sunt înregistrate peste 6.000 de evaluări. De interes îndeosebi pentru următoarele direcţii aplicative ale psihologiei: psihologie clinică, consiliere psihologică şi psihoterapie.

Campbell Collaboration – www.campbellcollaboration.org

Baza de date este inspirată după modelul de succes al Cochrane Collaboration, cu scopul de a colecta şi evalua toate studiile de tip experimental sau cvasiexperimental din domenii variate ale ştiinţelor sociale. Este vorba de intervenţii aplicate în domenii precum educaţie, asistenţă socială, politici publice, justiţie şi criminalitate. La ora actuală există peste 200 de sinteze realizate. De interes îndeosebi pentru direcţiile aplicative: psihologie şcolară, psihologie comunitară, psihologie juridică, psihologie economică, psihologie socială aplicată.

Bibliografie

APA (2012). *Society for Clinical Psychology. American Psychological Association. Division 12*. Disponibilă la adresa http://www.div12.org/PsychologicalTreatments/faq.html.

APA Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *American Psychologist*, 61, pp. 271-285.

Atkins, D., Best, D., Briss, P.A., Eccles, M., Falck-Ytter, Y., Flottorp, S., Guyatt, G.H. *et al.* (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal*, 328 (7454), p. 1490.

Balshem, H., Helfand, M., Schunemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., Vist, G.E. *et al.* (2011). GRADE guidelines 3: rating the quality of evidence – introduction. *Journal of Clinical Epidemiology*, 64, pp. 401-406.

Bennett, S., Tooth, L., McKenna, K., Rodger, S., Strong, J., Ziviani, J. *et al.* (2003). Perceptions of evidence-based practice: a survey of Australian occupational therapists. *Australian Occupational Therapy Journal*, 50, pp. 13-22.

Bîrle, D. (2010). *Implicaţii psihologice ale alegerii carierei*. Teză de doctorat, Universitatea „Babeş-Bolyai”, Cluj-Napoca.

Briner, R.B., Rousseau, D.M. (2011). Evidence-based I-O psychology: Not there yet. *Industrial and Organizational Psychology*, 4, pp. 3-22.

Campbell Collaboration (2000). *The Campbell Collaboration. What helps? What harms? Based on what evidence*. Disponibil la adresa: http://www.campbellcollaboration.org.

Chambless, D.L., Baker, M.J., Baucom, D.H., Beutler, L.E., Calhoun, K.S., Crits-Christoph, P., Daiuto, A. *et al.* (1998). Update on empirically validated therapies, II. *Clinical Psychologist*, 51, pp. 3-16.

Chambless, D.L., Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, pp. 7-18.

Chambless, D.L., Ollendick, T.H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52, pp. 685-716.

Chambless, D.L., Sanderson, W.C., Shoham, V., Bennett Johnson, S., Pope, K.S., Crits-Christoph, P., Baker, M. *et al.* (1996). An update on empirically validated therapies. *Clinical Psychologist*, 49, pp. 5-18.

Cochrane Collaboration (1993). *The Cochrane Collaboration. Working together to provide best evidence for health care*. Disponibil la adresa: http://www.cochrane.org.

Cohen, L.L., La Greca, A.M., Blount, R.L., Kazak, A.E., Holmbeck, G.N., Lemanek, K.L. (2008). Introduction to special issue: Evidence-based assessment in pediatric psychology. *Journal of pediatric psychology*, 33, pp. 911-915.

Craig, J.V., Smyth, R.L. (2002). *The evidence-based practice manual for nurses*. New York: Elsevier Health Sciences.

David, D. (2006). *Metodologia cercetării clinice*. Iaşi: Editura Polirom.

David, D., Montgomery, G.H. (2011). The scientific status of psychotherapies: A new evaluative framework for evidence-based psychosocial interventions. *Clinical Psychology: Science and Practice*, 18, pp. 89-99.

Gaudiano, B.A., Statler, M.A. (2001). The scientist practitioner gap and graduate education: Integrating perspectives and looking forward. *The Clinical Psychologist*, 54, pp. 12-18.

Gibbs, L. (2002). *Evidence-based practice for the helping professions*. Belmont, CA: Wadsworth Publishing, Thomson Higher Education.

Guyatt, G., Oxman, A.D., Akl, E., Kunz, R., Vist, G., Brozek, J., Norris, S. *et al.* (2011). GRADE guidelines 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64, pp. 383-394.

Humphris, D., Littlejohns, P., Victor, C., O’Halloran, P., Peacock, J. (2000). Implementing evidence based practice: factors that influence the use of research evidence by occupational therapists. *British Journal of Occupational Therapy*, 63, pp. 516-522.

Ig, R., Kwan, I. (2008). School-based driver education for the prevention of traffic crashes. *The Cochrane Library*, 4, disponibil la adresa: http://www.thecochranelibrary.com.

Kratochwill, T.R. (2012). *Task force on evidence-based interventions in school psychology. Procedural and coding manual for review of evidence-based interventions*. Disponibil la adresa: http://www.indiana.edu/~ebi/EBI-Manual.pdf.

Kratochwill, T.R., Shernoff, E.S. (2004). Evidence-based practice: promoting evidence-based interventions in school psychology. *School Psychology Review*, 33, pp. 34-48.

Kratochwill, T.R., Stoiber, K.C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the Procedural and Coding Manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly*, 17, pp. 341-389.

Ledford Jr, G.E., Heneman III, H.G. (2011). Skill-based pay. *The SIOP/SHRM Library*, disponibil la adresa: http://www.siop.org/siop-shrm/default.aspx.

Levant, R.F., Hasan, N.T. (2008). Evidence-based practice in psychology. *Professional Psychology: Research and Practice*, 39, pp. 658-662.

Lissiman, E., Bhasale, A.L., Cohen, M. (2009). Garlic for the common cold. *The Cochrane Library*, disponibil la adresa: http://www.thecochranelibrary.com.

Liţă, S., Stan, A., Sava, F.A. (2008). The development and use of Romanian system for reviewing and certification of psychological tests. Lucrare prezentată la cel de-al XXIX-lea Congres Internaţional de Psihologie, Iulie 20-25, Berlin, Germania.

NICE (2005). *National Institute for Health and Clinical Excellence*. Disponibil la adresa: http://www. nice.org.uk.

Noonan, E. (2009). Evidence-based pedagogy. *The Campbell Library*, disponibil la adresa: http://www.campbellcollaboration.org.

Norcross, J.C., Beutler, L.E., Levant, R.F. (eds.) (2006). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. APA, Washington, DC.

Overholser, J.C. (2010). Ten criteria to qualify as a scientist-practitioner in clinical psychology: An immodest proposal for objective standards. *Journal of Contemporary Psychotherapy*, 40, pp. 51-59.

Peterson, K. (2006). Practice-based primary care research-translating research into practice through advanced technology. *Family Practice*, 23, pp. 149-150.

Petrosino, A., Turpin-Petrosino, C., Buehler, J. (2009). *„Scared straight" and other juvenile awareness programmes for preventing juvenile delinquency*. *The Cochrane Library*, disponibil la adresa: http://www.thecochranelibrary.com.

Rousseau, D.M. (2006). Is there such a thing as evidence-based management?. *Academy of Management Review*, 31, pp. 256-269.

Rynes, S.L., Giluk, T.L., Brown, K.G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, 50, pp. 987-1008.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., Richardson, W.S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312, pp. 71-72.

Sackett, D.L., Strauss, S.E., Richardson, W.S., Rosenberg, W., Haynes, R.B. (2000). *Evidence-based medicine: How to practice and teach EBM* (ediţia a II-a). New York: Churchill Livingstone.

Schneider, B., Barbera, K.M. (2011). Driving customer satisfaction through HR: creating and maintaining a service climate. *The SIOP/SHRM Library*, disponibil la adresa: http://www.siop.org/siop-shrm/default.aspx.

Seligman, M.E.P. (1995). The effectiveness of psychotherapy. The consumer reports study. *American Psychologist*, 50, pp. 965-974.

Shernoff, E.S., Kratochwill, T.R., Stoiber, K.C. (2003). Training in evidence-based interventions (EBIs): What are school psychology programs teaching?. *Journal of School Psychology*, 41, pp. 467-483.

SIOP (2003). *Principles for the validation and use of personnel selection procedures* (ediţia a IV-a). Disponibil la adresa: http://www.siop.org/_Principles/principles.pdf.

SIOP/SHRM Collaboration (2011). *SIOP and SHRM: promoting evidence-based HR*. Disponibil la adresa http://www.siop.org/siop-shrm/default.aspx.

Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters, what you need to know. *Social Science Research*, 63, pp. 611-631.

Stricker, G. (1997). Are science and practice commensurable?. *American Psychologist*, 52, pp. 442-448.

Vîrgă, D. (2007). *Psihologie experimentală – de la teorie la practică*. Timişoara: Editura Universităţii de Vest.

Wikipedia (2012). List of academic databases and search engines. Disponibil la adresa: http://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines.

Wilson, J. L., Armoutliev, E., Yakunina, E., Werth, J.L. (2009). Practicing psychologists' reflections on evidence-based practice in psychology. *Professional Psychology: Research and Practice*, 40, pp. 403-409.

Wilson, S.J., Tanner-Smith, E.E., Lipsey, M.W., Steinka-Fry, K., Morrison, J. (2011). Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *The Campbell Library*, disponibil la adresa: http://www.campbellcollaboration.org.

1. O altă traducere, ce va fi utilizată alternativ pentru EBP, este aceea de practică bazată pe dovezi empirice.

2. În acest sens ar merita analizate în contrast normativele practicii EBP cu situaţia întâlnită în multe cabinete de psihologia muncii şi transporturilor din România. Din nefericire, în cele mai multe cazuri, deciziile de acordare a avizului psihologic (apt/inapt) au fost luate în baza unor metodologii lipsite complet de orice validare ştiinţifică (cu instrumente psihologice depăşite şi/sau cu proprietăţi psihometrice discutabile; în absenţa studiilor de validare a eficienţei procesului de testare etc.). Aceste neajunsuri au condus în desuetudine practica din cabinetele

de profil şi a contribuit, alături de alţi factori politici, la amendarea legii 355 din 2007 la finalul anului 2011, soldată cu eliminarea obligativităţii examinării psihologice periodice în cele mai multe situaţii.

3. Accesul este gratuit la sintezele din *Campbell Collaboration*, respectiv din SIOP/SHRM Collaboration, atât pentru rezumate, cât şi pentru rapoartele integrale. Accesul la sintezele din *Cochrane Collaboration* este gratuit numai pentru rezumate. Pentru rapoartele integrale accesul se face contra cost cu unele excepţii (gratuitate pentru utilizatorii din UK sau în baza unor donaţii).

4. Din punct de vedere istoric, arhiva *Cochrane Collaboration* a fost iniţiată în anul 1993, în UK (denumirea fiind dată în cinstea lui Archie Cochrane [medic], care a fost primul cercetător care a atras atenţia la începutul anilor ’70 asupra unei probleme extinse, aceea de a utiliza în practică intervenţii despre care nu se ştie nimic cu privire la nivelul lor de eficienţă – cât de benefice sunt acestea). Arhiva *Campbell Collaboration* a fost demarată mai târziu, în anul 2000, denumirea fiind dată în cinstea psihologului Donald Campbell care a avut contribuţii importante metodologice, în special în identificarea distorsiunilor care pot afecta concluziile studiilor, îndeosebi a celor de tip cvasiexperimental.

5. Totuşi, în lipsa accesului la rapoarte în format integral, în cazul *Cochrane Collaboration* este dificil să se răspundă din perspectiva EBP la întrebarea legată de aplicabilitate (vezi tabelul 2.2).

6. Ideea unei liste de servicii avizate pe baza nivelului lor de validitate ştiinţifică a fost preluată de către Comisia Metodologică a Colegiului Psihologilor din România cu referire la avizarea testelor (www.copsi.ro). Astfel, unele teste au primit aviz pe perioadă nedeterminată (cele cu nivelul cel mai ridicat de dovezi privind existenţa unor proprietăţi psihometrice solide), iar altele au primit un aviz pe perioadă determinată (cele cu dovezi minimal acceptabile din perspectivă psihometrică). Desigur, există şi varianta unui aviz nefavorabil (pentru probele psihologice care nu îndeplinesc nici măcar criteriile minimale). Din păcate, lista oficială prezintă doar probele din primele două categorii, astfel că nu este clar dacă o probă a fost omisă din listă pentru că a fost respinsă sau pentru că nu a fost supusă procesului de evaluare.

7. Pentru o listă extinsă a bazelor de date internaţionale existente şi a domeniului principal de aplicabilitate recomand consultarea Wikipedia (2012).

8. În domeniul psihologiei organizaţionale există mult mai puţine resurse de acest gen (vezi, spre exemplu, ghidul dezvoltat pe tema selecţiei de personal – SIOP, 2003). De altfel, SIOP (*Society of Industrial and Organizational Psychology*) este principalul promotor al modelului EBP în domeniul psihologiei muncii şi organizaţională. În această direcţie, începând cu 2011 sub egida SIOP, împreună cu SHRM (*Society for Human Resources Management*), au început să fie publicate sinteze utile pentru practicieni pe baza informaţiilor validate ştiinţific (vezi sintezele realizate de Ledford Jr. şi Heneman III, 2011; Schneider şi Barbera, 2011). Totuşi, aceste sinteze au o structură apropiată mai degrabă de evaluările narative decât de cele sistematice. Acest demers de colaborare poartă denumirea în limba engleză de SIOP/SHRM Collaborative

Series şi are drept scop promovarea modelului bazat pe dovezi empirice valide ştiinţific în domeniul resurselor umane.

9. Se au în vedere aspecte care ţin de validitatea internă. Desigur, pot fi generate şi alte tipuri de piramide care să aibă în vedere alte tipuri de validitate, cum ar fi validitatea externă. În astfel de situaţii, sondajele de opinie sau panelurile efectuate pe un eşantion reprezentativ de persoane ar ocupa locurile dinspre vârful piramidei, chiar dacă ar fi studii de tip non-experimental.

10. Din cauza eterogenităţii ridicate a calităţii designurilor studiilor cuprinse într-o metaanaliză de studii experimentale, există tendinţa întâlnită în domeniul medical de a plasa în topul încrederii rezultatele obţinute în cadrul unui studiu randomizat controlat extins (*large-scale randomized trials*) (David, comunicare personală, 22 aprilie 2012). Acest tip de studiu experimental are trei caracteristici de bază: (a) implementarea se face prin cooperarea unor echipe de cercetare independente, în locaţii independente; (b) se realizează pe un volum impresionant de participanţi (de obicei N > 1.000) şi (c) respectă un nivel ridicat de exigenţă din perspectivă metodologică. Având în vedere aceste caracteristici tehnice şi a resurselor uriaşe pe care le implică, acest gen de design este, după cunoştinţele mele, deocamdată absent din domeniul psihologiei.

11. În psihologie, prin tratament placebo se înţelege o intervenţie aparent similară cu cea din grupul experimental, dar care nu urmăreşte să modifice variabila dependentă de interes, efectul intervenţiei placebo fiind localizat asupra unui aspect nerelaţionat cu variabila dependentă. În cazul efectului placebo, ideal ar fi ca participanţii la studiu să nu poată distinge clar ce intervenţie au urmat.

12. Iată un exemplu de intervenţie placebo (Bîrle, 2010) într-un studiu de testare a eficienţei unui program de training menit să îmbunătăţească capacitatea de decizie privind cariera şcolară la elevii aflaţi în an terminal de liceu: atât participanţii din grupul experimental, cât şi cei din grupul placebo au urmat 10 sesiuni de training a câte 50 de minute fiecare. Ceea ce a diferit a fost conţinutul predat. În cazul celor din grupul placebo, acesta nu avea legătură directă cu abilitatea de a lua decizii privind cariera, fiind predate informaţii cu caracter general despre comunicare şi rezolvarea de conflicte.

13. În opinia mea, sistemul GRADE este cel mai apropiat de realitatea practicienilor, atât prin simplitate, cât şi prin faptul că ţine seama de aspecte importante pentru practica de zi cu zi, cum ar fi analiza similarităţii dintre situaţia de rezolvat (caracteristicile clientului) şi cele ale cercetărilor în care au fost obţinute rezultatele.

14. Seligman (1995) identifică cinci limite majore inerente studiilor experimentale descriptiv-cauzale (*efficacy*) prin raportare la practica cotidiană: (a) experimentele au o durată standard limitată spre deosebire de practică unde durata şi intensitatea programului de intervenţie variază; (b) practicienii pot interveni pe parcursul derulării intervenţiei corectând sau modificând unele aspecte gândite iniţial, în timp ce în experimentele de acest tip terapeuţii urmează strict recomandările din manualul de intervenţie; (c) în realitate clienţii îşi aleg de

multe ori forma de psihoterapie urmată în baza preferinţelor personale, iar această motivaţie, absentă în studiile bazate pe randomizarea participanţilor, poate augmenta efectul obţinut; (d) în practică cei mai mulţi clienţi au multiple comorbidităţii (mai multe tulburări), în timp ce în experimentele de tip *efficacy* sunt excluşi asemenea indivizi; (e) în practică este relevantă funcţionarea per ansamblu a individului, în timp ce în experimentele menţionate relevantă este eficienţa intervenţiei asupra variabilei dependente ţintă (*primary outcome*).

15. În domeniul organizaţional au fost efectuate foarte puţine studii experimentale (randomizate), deoarece majoritatea intervenţiilor au fost implementate şi finanţate direct de către beneficiar (adesea organizaţii orientate spre profit). Din acest motiv cele mai multe studii din zona testării eficienţei unei intervenţii sunt de tip cvasiexperimental. Prin urmare, raportarea ar trebui să ţină seama de acest specific.

Capitolul 3
Strategii de cercetare în psihologie

Scopul principal al acestui capitol este acela de oferi cititorilor o grilă pragmatică de lectură a aspectelor ce ţin de metodologia cercetării, pe care le consider esenţiale pentru dezvoltarea unei priviri de ansamblu despre activitatea de cercetare din domeniul psihologiei. Prin urmare, în acest capitol nu ne vom opri asupra fundamentelor activităţii ştiinţifice, asupra noţiunilor de epistemologie sau asupra tradiţiilor şi curentelor de abordare din filosofia ştiinţei şi din logică. Cei interesaţi de aceste aspecte conceptuale bazale pot găsi informaţii relevante în primele două capitole din David (2006). În schimb, scopul acestui capitol este acela de a contribui la dezvoltarea gândirii critice şi a spiritului analitic, atunci când se planifică un studiu psihologic, respectiv atunci când se analizează critic rezultatele provenite din cercetări realizate de alţii.

Pentru a atinge acest deziderat, am împărţit capitolul de faţă în şase secţiuni aparent eterogene, din combinarea cărora sperăm să oferim cititorilor interesaţi acea grilă de lectură critică utilă şi o viziune de ansamblu asupra demersului de cercetare. Accentul este pus pe metodologia de cercetare de tip cantitativ, aceasta fiind perspectiva dominantă la ora actuală în psihologie.

3.1. False dileme în cercetarea psihologică

Activitatea de cercetare presupune o serie de decizii din partea celui care o efectuează, fiecare dintre acestea având un impact important asupra rezultatului final obţinut. Între aceste decizii de luat se regăsesc şi o serie de false dileme sau afirmaţii. Pe cele mai răspândite dintre acestea, întâlnite în cei peste 10 ani de activitate în zona psihologiei academice, le vom prezenta în continuare.

3.1.1. Metode de cercetare cantitative vs metode de cercetare calitative în psihologie

Pe de o parte, am întâlnit colegi care susţin cu tărie că cea mai potrivită modalitate de a accesa nivelul de profunzime al psihicului este prin metode calitative de cercetare, deoarece metodele cantitative tratează la un nivel superficial fenomenele psihice. Am sesizat că această poziţie nefavorabilă metodologiei cantitative este îmbrăţişată îndeosebi de cei care manifestă o nesiguranţă în stăpânirea aparatului metodologic şi/sau statistic, ceea ce mă face să cred că rolul acestor credinţe este unul defensiv, acţionând ca un veritabil mecanism de apărare. Pe de altă parte, am întâlnit colegi de breaslă care încercau să dezlipească de sens orice demers de cercetare psihologică, înlocuindu-l cu un şir lung de cifre (rezultate statistice), desprinse de semnificaţia lor teoretică/psihologică. Pentru aceştia, cercetarea de tip calitativ are doar un rol secundar, acela de a oferi date congruente pentru a confirma sau ilustra rezultatele cantitative obţinute.

Din punctul meu de vedere, o asemenea dilemă, cantitativ *vs* calitativ, este una falsă, deoarece problema superiorităţii unei forme sau alta de cercetare nu se poate pune în termeni abstracţi, ci numai în legătură cu un anumit obiectiv de studiat. Aşadar, problema pusă în termeni corecţi ar fi

aceea de a stabili care dintre cele două forme de metodologie este mai potrivită pentru un anumit obiectiv de cercetare vizat. De pildă, cercetarea calitativă pare a fi mai potrivită atunci când obiectivul studiului vizează înţelegerea unor fenomene din perspectiva actorului, descriind realitatea aşa cum este ea percepută de către participanţi, pornind de la premisa că este mai important ceea ce percep oamenii ca fiind realitate, decât realitatea observabilă însăşi (Kvale *apud* Gilbert şi Irons, 2008). În schimb, metodologia de tip cantitativ este mai potrivită atunci când obiectivul studiului este de a stabili existenţa unei relaţii de tip cauză-efect, cum ar fi identificarea măsurii în care o anumită formă de psihoterapie este eficientă în combaterea atacurilor de panică.

3.1.2. O cercetare este cu atât mai valoroasă cu cât analizele statistice sunt mai complexe

Aparent, apelul la metode statistice sofisticate, cum ar fi modelarea prin ecuaţii structurale, apelul la regresii logistice sau analiza de clustere, oferă o prestanţă superioară oricărui demers ştiinţific. Într-o anumită măsură, o asemenea concluzie este sprijinită de multitudinea şi complexitatea relaţiilor dintre diverse fenomene. Realitatea poate fi surprinsă mai bine prin tehnici statistice multivariate decât prin cele aparţinând statisticii uni- sau bivariate. În esenţă, însă, calitatea ridicată a unei cercetări este determinată îndeosebi de problema investigată şi de noutatea (importanţa) teoretică sau aplicativă a subiectului studiat. A apela la ecuaţii structurale pentru a demonstra lucruri arhicunoscute (de exemplu, că inteligenţa şi motivaţia şcolară, chiar surprinse ca variabile latente, prezic performanţa şcolară a elevilor) reprezintă un demers ştiinţific desuet, chiar dacă corect din punct de vedere statistic. De asemenea, apelul la tehnici statistice

complexe pentru a compara empiric modele care nu au nicio relevanţă aplicativă sau nicio fundamentare teoretică reprezintă o altă tendinţă care trebuie descurajată.

În consecinţă, o prelucrare statistică de excepţie, deşi necesară, nu ţine locul unei cercetări de o calitate superioară. Ceea ce contează, în primul rând, este importanţa problemei investigate şi adecvarea metodologiei de cercetare (design, metode de cercetare şi de analiză a datelor) la obiectivul propus al studiului.

3.1.3. O cercetare este mai valoroasă dacă îşi propune să investigheze mai multe aspecte

Gândirea de tipul „mai mult este mai bine” am regăsit-o destul de frecvent, mai ales în rândul studenţilor, atunci când aceştia îşi puneau problema în studiile lor, dacă trei ipoteze sau patru probe psihologice utilizate sunt suficiente, dacă două obiective ajung etc. Desigur, asemenea dileme constituie un indiciu potrivit pentru a detecta lipsa abilităţii de a face cercetare, prin neînţelegerea aspectului central situat la baza activităţii de cercetare. Astfel, orice demers de cercetare trebuie să se raporteze în permanenţă la întrebarea supusă investigării (obiectivul cercetării). Numai prin raportare la acest aspect esenţial se pot formula adecvat ipoteze de cercetare şi se pot stabili corect aspecte specifice, precum tipul de design al cercetării, numărul de instrumente necesare (variabile de operaţionalizat) etc.

În consecinţă, numărul de obiective, de ipoteze, de instrumente utilizate nu este relevant în sine, ci se stabileşte în funcţie de aspectele investigate în cercetarea respectivă. Pot exista cercetări cu un singur obiectiv şi o singură ipoteză principală de investigat, mult mai valoroase

din punct de vedere teoretic sau aplicativ decât cercetări care investighează simultan mai multe aspecte redundante/arhicunoscute (cum ar fi asocierea dintre nivelul de anxietate şi cel de neuroticism) sau irelevante pentru obiectivul studiului (cum ar fi diferenţa dintre băieţi şi fete cu privire la numărul de ore de vizionare zilnică la TV într-un studiu care urmăreşte să stabilească dacă există o asociere între stilul parental autoritar şi comportamentul prosocial al copiilor).

3.1.4. La nivelul studiilor doctorale în psihologie, cercetarea derulată trebuie să propună ceva original, care să revoluţioneze domeniul de interes al tezei

Ideea că ştiinţa progresează prin experimente unice, cruciale este o percepţie predominant greşită, determinată mai degrabă de titlurile din mass-media şi de absenţa unei culturi ştiinţifice, inclusiv în rândul multor absolvenţi de studii superioare. Fără a nega posibilitatea unor asemenea excepţii, cel mai frecvent cunoaşterea progresează gradual prin contribuţii limitate la problematica investigată. Deoarece unul dintre cele mai importante mesaje ale acestei lucrări este că nicio cercetare nu este perfectă din perspectiva validităţii concluziilor rezultate (vezi detalii în secţiunea 3.4 a acestui capitol), este important ca orice demers investigativ să fie conectat la cele anterioare şi să încerce să rezolve o necunoscută pe care studiile anterioare fie nu au abordat-o, fie nu au reuşit să o rezolve în mod concludent. Cu alte cuvinte, un demers respectabil pentru orice teză de doctorat, respectiv pentru orice lucrare de cercetare în genere, este acela de a conduce unul sau mai multe studii care

să elimine unele dintre neajunsurile sesizate în cercetările anterioare efectuate pe tematica respectivă. Desigur, acest demers trebuie însoţit de sublinierea impactului teoretic şi/sau aplicativ al studiilor propuse. Pentru a ilustra acest principiu al progresului gradual în cunoaştere, citiţi conţinutul casetei 4.2 din capitolul următor (capitolul 4), în care sunt prezentate o serie de studii succesive cu privire la relaţia cauzală dintre dispoziţia afectivă pozitivă şi nivelul de altruism.

3.2. Probleme stringente în cercetarea psihologică din România

Dacă aspectele amintite anterior reprezintă false probleme cu care se confruntă mulţi consumatori sau producători de cercetare din România, voi aminti în cele ce urmează patru probleme importante care pot impieta dezvoltarea sănătoasă a cercetării în psihologie din ţară.

3.2.1. Cultura ştiinţifică deficitară

Există cercetări bune şi cercetări cu rezultate distorsionate. Din nefericire însă, am constatat că o bună parte dintre studenţii sau absolvenţii de psihologie nu pot distinge corect între diferitele tipuri de designuri de cercetare (experimental, cvasiexperimental sau non-experimental), fiindu-le cu atât mai greu să facă diferenţe între cercetări aparţinând aceleiaşi clase de design din perspectiva riscurilor existente în cazul validităţii concluziilor studiului. În lipsa acestei competenţe tehnice, ei tind să acorde credit informaţiilor prezentate în studiile respective, indiferent de gradul acestora de acoperire. În spatele acestui credit stă premisa uneori

eronată că studiile prezintă adevăruri valide, de vreme ce ele au fost publicate şi sunt efectuate de către personalităţi din domeniul academic sau de formatori de şcoli psihoterapeutice etc. Însă nu toate studiile ajung să fie publicate în reviste prestigioase, fundamentate pe un proces de evaluare colegială foarte exigent. O parte semnificativă din studii ajung să fie publicate în reviste cu o politică editorială mai puţin riguroasă, aspect decelat cu dificultate de către nespecialişti. Acest segment de consumatori de literatură academică, posesor al unei culturi metodologice deficitare, este cel mai predispus să cadă în capcanele abordărilor pseudoştiinţifice reliefate pe larg în primul capitol al cărţii, cum ar fi îndeosebi extinderea fără temei a concluziilor, distorsiunea de confirmare, inversarea demersului de validare ştiinţifică ş.a.m.d.

3.2.2. Probleme etice în derularea studiilor şi analiza rezultatelor

Spre deosebire de situaţia existentă în statele occidentale, aspectele etice relaţionate activităţii de cercetare psihologică din România rămân în multe cazuri abordate doar la un nivel teoretic. La această stare de fapt contribuie deopotrivă absenţa unei instanţe din multe universităţi care să evalueze specific studiile de cercetare propuse încă din stadiul de proiect, cât şi lipsa unei cutume din partea cercetătorilor de a utiliza formulare de consimţire informată atunci când îşi derulează cercetările pe diferite categorii de participanţi[1]. O problemă la fel de gravă, întâlnită atât în rândul practicienilor psihologi, dar şi în unele medii studenţeşti, este aceea a utilizării de instrumente psihologice fără a avea drept de utilizare (cu încălcarea drepturilor de autor asupra acestora).

Nu în ultimul rând, o altă problemă de natură etică, adesea tratată cu tăcere, este aceea a falsificării/fabricării rezultatelor sau a utilizării selective a datelor provenite din analizele statistice efectuate (ignorate fiind cele care nu sprijină ipotezele formulate şi/sau rezultatele nesemnificative statistic). O discuţie elaborată asupra acestei teme este prezentată în capitolul 8, secţiunea 8.2.2. Pentru moment ne vom rezuma doar să amintim cele trei mari fraude existente: (a) fabricarea datelor; (b) falsificarea datelor; şi (c) plagiatul pentru care există sancţiuni drastice care pot merge până la retragerea titlurilor ştiinţifice sau profesionale dobândite. Restul aspectelor constituie abateri de la buna conduită care trebuie corectate de autor atunci când sunt semnalate.

3.2.3. Probleme metodologice I: inadaptarea literaturii de specialitate la obiectivele studiului

O afirmaţie dură, dar din nefericire nu foarte îndepărtată de adevăr, este aceea că, în foarte multe cazuri, ceea ce apare în secţiunile introductive ale unor articole publicate în reviste româneşti de specialitate poate fi inclus în categoria „maculatură”. Adesea, secţiunile introductive se apropie de ideea unor referate despre conceptele uzitate în lucrare. De exemplu, într-un studiu care vizează rolul predictiv al inteligenţei generale asupra performanţei în muncă a unui lot de cadre didactice mi-a fost dat să văd că nu se discută direct problema abordată („ce anume nu e clar asupra acestei relaţii din literatura anterioară?”, „de ce ar fi important să studiem această problemă?”, respectiv „ce şi cum anume va reuşi studiul nostru să rezolve o parte din necunoscutele semnalate?”). În schimb este posibil să întâlnim un întreg pomelnic despre istoricul evoluţiei conceptului de inteligenţă (de la Galton şi Binet până în

prezent), să găsim o prezentare detaliată a conceptului de măsurare a performanţei în muncă şi a tipologiilor asociate existente, plus, eventual, un scurt sumar necritic al studiilor care au abordat relaţia dintre inteligenţă şi performanţa profesională.

Cititorii interesaţi să-şi îmbunătăţească abilitatea de a redacta secţiunile introductive ale articolelor lor de specialitate au la dispoziţie un ghid prescriptiv pe această problematică în secţiunea 8.1.4 (capitolul 8).

3.2.4. Probleme metodologice II: inadaptarea metodologiei cercetării la obiectivele studiului

O altă problemă întâlnită destul de frecvent constă în utilizarea unui design de cercetare inadecvat pentru a răspunde specificului întrebărilor din studiu. Astfel, am întâlnit studii ce conţineau afirmaţii care sprijină existenţa unor relaţii de tip cauză-efect, deşi designul utilizat nu le permitea acest lucru, fiind vorba de studii corelaţionale. În mod similar, am sesizat situaţii în care se ofereau explicaţii (era postulat un anumit mecanism cauzal) în studii al căror design experimental nu permitea identificarea unor mecanisme ale schimbării etc.

Concluziile rezultate din asemenea studii conţin afirmaţii hazardate, ce nu se susţin din cauza lipsei controlului adecvat asupra factorilor de risc ce pun în pericol validitatea studiilor respective. În restul capitolului de faţă, accentul este pus pe reducerea acestui tip specific de erori, astfel încât să fie clar ce tip de studii şi ce fel de design metodologic aferent este mai potrivit pentru a răspunde la o anumită problemă de cercetat. În acest sens vom trece, mai întâi, în revistă tipurile majore de cercetare şi impactul lor asupra întrebărilor de cercetare care pot fi adresate.

3.3. Tipuri de cercetări şi relevanţa lor pentru procesul cunoaşterii

Prin cercetare în sens larg se înţelege o activitate care produce cunoaştere. Cercetarea ştiinţifică este un anumit tip de cercetare, deosebită de cea filosofică sau de cea teologică prin metodologia utilizată pentru a ajunge la „adevăr”, la acele cunoştinţe necesare rezolvării unor probleme (David, 2006). Din acest punct de vedere, al impactului pe care îl are în cunoaştere şi al sferei de problematici ce pot fi rezolvate, cercetarea ştiinţifică poate fi clasificată ca: (a) cercetare fundamentală; (b) cercetare aplicată; (c) cercetare pentru dezvoltare şi inovare.

În cadrul psihologiei şi a ştiinţelor sociale în general, cercetarea aspectelor psihice poate accentua fie realitatea observabilă, dată (independentă de credinţele celui evaluat), fie realitatea construită de subiectul cercetării (semnificaţia şi interpretarea realităţii construite de către subiect). În primul caz vorbim de (a) o paradigmă realistă, pozitivistă de abordare a psihologiei, iar în cel de-al doilea despre (b) o paradigmă constructivistă (hermeneutică sau fenomenologică).

Luând în calcul caracteristicile metodologiei de cercetare utilizate, cercetarea ştiinţifică poate apela la: (a) metodologii calitative sau la (b) metodologii cantitative. Tot din punct de vedere metodologic, luând în calcul gradul de control al variabilelor confundate vorbim despre: (a) studii experimentale; (b) studii cvasiexperimentale şi (c) studii non-experimentale.

În sfârşit, după obiectivul pe care îl poate avea un studiu şi pornind de la rolul cunoaşterii ştiinţifice de a descrie, prezice şi explica fenomene psihice, putem împărţi cercetările ştiinţifice în: (a) cercetări exploratorii; (b) cercetări descriptiv-analitice; (c) cercetări cauzale. Această ultimă

categorie poate fi la rândul ei divizată în cercetări cauzal-descriptive şi cercetări cauzal-explicative.

În cele ce urmează vom oferi o serie de informaţii suplimentare despre fiecare tip de clasificare, accentuând acele aspecte care evidenţiază tipul de întrebări de cercetare la care se poate răspunde printr-un anumit tip de cercetare. În acest demers, ne vom focaliza îndeosebi pe ultima dintre clasificări, deoarece aceasta adresează direct problematica selectării celui mai potrivit tip de design de cercetare în funcţie de întrebarea cercetării (obiectivul studiului).

3.3.1. Cercetare fundamentală vs cercetare aplicată vs cercetare pentru dezvoltare-inovare

La originea distincţiei dintre cele trei forme de cercetare se află conflictul dintre modelul cartezian de a vedea ştiinţa (rolul fundamental al ştiinţei este de a cunoaşte şi înţelege lumea) şi modelul lui Francis Bacon (rolul ştiinţei este de a ameliora bunăstarea oamenilor, utilitatea descoperirilor plasându-se înaintea cunoaşterii în sine) (Bevan, 1980 *apud* Kantowitz, Roediger III şi Elmes, 2009). Dacă ar fi să ne raportăm la cazul specific al psihologiei, cercetarea fundamentală se încadrează în primul model de cunoaştere, fiind interesată îndeosebi să găsească explicaţii pentru modul de funcţionare a psihicului uman (de exemplu, „cum poţi reduce un comportament indezirabil, în general?”), în timp ce cercetarea aplicată şi cea orientată spre dezvoltare-inovare se centrează asupra măsurii în care anumite cunoştinţe rezultate din cercetarea fundamentală pot fi aplicate în contexte specifice pentru a rezolva probleme concrete sau pentru a veni în sprijinul unor nevoi pe care le au oamenii (de exemplu, „cum poţi să

modifici comportamentul copiilor care suferă de enurezis, în direcţia reducerii acestor comportamente?”).

Din punct de vedere pedagogic, cele două tipuri majore de cercetare, fundamentală *vs* aplicată, se diferenţiază prin: (a) gradul de generalizare a rezultatelor obţinute (transferabilitate ridicată în cercetarea fundamentală *vs* transferabilitate limitată în cercetarea aplicativă)[2]; (b) prin gradul de utilitate a rezultatelor obţinute (utilitate practică imediată absentă/redusă – cercetarea fundamentală; utilitate practică imediată ridicată – cercetarea aplicativă); (c) accentul pe aspecte teoretice în dauna practicii (accent pe aspecte teoretice, pe cunoştinţe declarative în cercetarea fundamentală; accent pe implicaţii practice, pe cunoştinţe procedurale în cercetarea aplicativă).

În realitate, distincţia dintre ele este mai degrabă una teoretică, deoarece cel mai adesea relaţia dintre ele este sub forma unui continuum care are la un capăt cercetarea fundamentală, iar la celălalt capăt cercetarea aplicativă. Pentru a ilustra această idee, voi apela la un exemplu din psihologia socială. Dacă obiectivul studiului este acela de a evidenţia caracteristicile procesărilor cognitive stereotipice cu referire la outgrup ne aflăm mai degrabă pe teritoriul cercetării fundamentale. Dacă ne propunem să oferim soluţii care să reducă conflictele interetnice prin reducerea stereotipurilor negative cu referire la membrii unui outgrup ne deplasăm pe acel continuum dinspre cercetarea fundamentală spre cea aplicată. În sfârşit, dacă ne propunem să studiem efectul unei intervenţii de reducere a stereotipurilor negative faţă de rromi (bazându-ne pe informaţiile oferite de cercetările anterioare menţionate) ne aflăm în zona cercetării aplicate.

Din cadrul cercetării aplicate se desprinde un tip aparte de studii, constituite sub denumirea generică de cercetare pentru dezvoltare şi

inovare. Aceasta apare atunci când cercetarea aplicată atinge acel nivel care-i permite să răspundă direct nevoilor unor clienţi prin oferirea de servicii, produse sau tehnologii inovative solicitate de către aceştia. Cel mai adesea, produsul final în asemenea cazuri nu se materializează prin publicaţii ştiinţifice, ca în cazul cercetării fundamentale sau aplicate, ci prin brevete şi/sau prototipuri tehnologice.

În sfârşit, o idee importantă este aceea că cele trei forme de cercetare nu diferă din perspectiva metodologiei cercetării uzitate. În toate cele trei cazuri putem găsi deopotrivă studii experimentale sau corelaţionale, studii cantitative sau calitative. Totuşi, în cercetarea fundamentală se observă un accent ridicat pus pe studiile de tip experimental, îndeosebi spre acele studii cauzal-explicative (vezi secţiunea 3.3.5) ce permit oferirea de explicaţii pentru relaţiile de tip cauză-efect constatate.

3.3.2. Cercetări în paradigma pozitivistă vs paradigma interpretativă

Pe lângă importanţa şi utilitatea practică a problemei cercetate, o amprentă importantă asupra cercetării derulate o are perspectiva filosofică din care este abordată tema respectivă. Ştiinţe naturale *vs* ştiinţe umaniste; etic *vs* emic; pozitivism *vs* hermeneutică (interpretivism); nometetic *vs* ideografic; cantitativ *vs* calitativ; naturalism *vs* relativism; dat *vs* construcţie; psihologia *hard* şi psihologia *soft* etc. reprezintă diferite terminologii utilizate pentru a sublinia o distincţie importantă existentă în cercetarea psihologică.

Pe de o parte se află cercetătorii susţinători ai paradigmei pozitiviste. Conform acestora, cunoaşterea ştiinţifică rezultă din aplicarea metodologiei ştiinţifice pentru a testa teoriile (adesea subiective)

formulate. Astfel, se accentuează ideea studiului riguros al comportamentului uman, pornind de la premisa că fenomenele psihice, la fel ca şi relaţiile sociale funcţionează în baza unor legităţi care trebuie descoperite, asemenea legilor (de exemplu, gravitaţia) din ştiinţele naturale. Prin urmare, accentul în această perspectivă cade asupra determinării suportului empiric existent pentru afirmaţiile făcute (de exemplu, surprinderii unei relaţii cauzale) şi pe controlul/eliminarea aspectelor introspective, subiective, considerate a pune în pericol validitatea studiilor. În această paradigmă, aspectele subiective nu sunt de interes decât din perspectiva măsurilor de control, în vederea eliminării riscului ca asemenea aspecte subiective să afecteze (distorsioneze) cunoaşterea fenomenelor studiate.

Pe de altă parte se află cercetătorii susţinători ai ideii că important pentru domeniul ştiinţelor sociale (psihologia fiind inclusă formal printre acestea) este să se studieze mai degrabă valorile, simbolurile, credinţele personale şi normele sociale, toate acestea accentuând construcţiile psihologice, perspectiva social istorică, subiectivă asupra lucrurilor, ţinându-se cont de contextul cercetării şi de dinamica interacţiunii dintre cercetător şi participanţi. În acest caz, accentul nu cade pe surprinderea unor relaţii deterministe, cauzale, specifice, ci pe surprinderea intenţiilor, a întregului, a modului în care participanţii interpretează şi înţeleg ceea ce trăiesc din punct de vedere psihologic, dar şi a influenţei pe care mediile social, istoric şi politic le au asupra acestor construcţii.

De asemenea, în abordarea interpretativă, aprecierea calităţii unei cercetări se face apelând la criterii diferite faţă de abordarea pozitivistă. În această abordare, demersul ştiinţific urmăreşte surprinderea nivelului de împărtăşire al unor opinii/trăiri între participanţi cu privire la problematica investigată, cu interpretarea semnificaţiei acestora în

gradiente diferite, în funcţie de paradigma dominantă de abordare (constructivistă, fenomenologică, hermeneutică, feministă, etnografică etc. sau a unei combinaţii dintre acestea). De exemplu, în paradigma fenomenologiei hermeneutice, accentul este pus pe găsirea unor semnificaţii (interpretarea) evenimentelor trăite, luând în calcul inclusiv detalii care pot oferi insighturi despre fenomenul analizat. Pentru a face acest lucru cercetătorul apelează la bagajul său de cunoştinţe, de pre-înţelegere a lumii, luând în calcul contextul social şi istoric în care trăiesc participanţii la studiu (Laverty, 2003).

Spre deosebire de această paradigmă, îmbrăţişată îndeosebi de ştiinţele umaniste, paradigma pozitivistă apelează la alte reguli de apreciere a calităţii unor studii. Dacă în abordarea de tip interpretativ cercetătorul introduce ca input cunoştinţele sale şi se bazează pe subiectivism pentru a desprinde semnificaţii, în paradigma pozitivistă cercetătorul încearcă să elimine orice input subiectiv din interpretarea datelor, bazându-se doar pe cel oferit din datele colectate prin apelul la o metodologie riguroasă. De aceea, în analiza calităţii cercetărilor efectuate, paradigma pozitivistă ţine cont de nivelul de întemeiere a concluziilor, din perspectiva valorii lor de adevăr („cât de siguri suntem că afirmaţia respectivă este adevărată sau falsă?”), acest lucru fiind determinat în funcţie de adecvarea metodologiei utilizate la întrebarea cercetării.

Această distincţie esenţială pentru aprecierea calităţii unei cercetări (prin prisma „valorii ei de adevăr obiectiv”, respectiv din prisma „progresului în înţelegerea fenomenului şi a nivelului de împărtăşire a esenţialului din fenomenul trăit”) este deseori ignorată, fapt ce poate duce la confuzii importante şi rezultate distorsionate. Un exemplu grăitor în acest sens este studiul tulburărilor emoţionale ale adolescenţilor rămaşi acasă şi ai căror părinţi muncesc în străinătate.

Apelând la o metodologie calitativă de investigare, specifică unui demers interpretativ, vom putea ajunge la o concluzie împărtăşită de participanţi, aceea a experimentării unor dificultăţi de ordin emoţional, resimţite subiectiv ca urmare a dorului de părinţi şi a insuficienţei suportului emoţional perceput. Dar este această metodologie capabilă să răspundă la o întrebare a cercetării de tip cauzal (plecarea părinţilor determină probleme emoţionale la copii)? Răspunsul este unul negativ. Întrebări care vizează relaţii de tip cauză-efect necesită apelul la abordări metodologice riguroase, focalizate pe controlul experimental astfel încât să permită stabilirea unei valori de adevăr (adevărat sau fals) cu privire la întrebarea din studiu (pentru mai multe detalii privind rigurozitatea controlului experimental, vezi secţiunea 4.1 din capitolul următor, îndeosebi prima parte). Pentru moment ne vom rezuma doar să amintim că pot exista o serie de variabile potenţial confundate pentru apariţia unor tulburări emoţionale la adolescenţii ai căror părinţi sunt plecaţi în străinătate (Sava, 2011), cum ar fi: (a) stilul parental autoritar mai des întâlnit în familii cu nivel scăzut de educaţie (status întâlnit frecvent printre familiile în care părinţii sunt plecaţi la muncă în străinătate) sau (b) relaţii de cuplu conflictuale existente încă dinaintea plecării în străinătate (mai des întâlnite în cazul acestui tip de familii decât în familiile cu ambii părinţi acasă). Astfel, este posibil ca eventualele tulburări emoţionale observate la adolescenţii ai căror părinţi sunt plecaţi la muncă în străinătate să nu fie cauzate de absenţa părinţilor, ci să fie existente încă dinaintea plecării lor, determinate de o vulnerabilitate emoţională crescută, specifică unor contexte familiale în care adolescentul este martor al unor conflicte repetate între părinţi, precum şi „beneficiar” al unui stil de educaţie autoritar, bazat predominant pe măsuri de constrângere.

În schimb, aceeaşi metodologie calitativă ar fi adecvată pentru a răspunde la o întrebare de cercetare care ar urmări să descrie „ce cred adolescenţii despre impactul emoţional determinat de plecarea părinţilor la muncă în străinătate”. În acest caz, miezul întrebării constă în decelarea semnificaţiilor şi interpretărilor pe care adolescenţii le oferă trăirilor afective experimentate în noul context familial – în absenţa părinţilor de acasă, indiferent dacă credinţele lor sunt conforme cu realitatea sau unele distorsionate[3].

Tabelul 3.1. *Sumar al particularităţilor cercetărilor în funcţie de paradigma filosofică utilizată*[4]

Aspect evaluat	Paradigma pozitivistă	Paradigma constructiv-hermeneutică
Asumpţii ontologice	Fenomenele psihosociale au legităţi ce pot fi descoperite, asemeni legităţilor din fizică.	Nu există un adevăr absolut, ci diferite realităţi subiective şi adevăruri.
Asumpţii epistemologice	Cunoaşterea este obiectivă şi poate fi atinsă prin metode ştiinţifice riguroase.	Cunoaşterea este subiectivă, dependentă cultural şi filtrată de percepţiile şi trăirile individului evaluat.
Rolul cercetătorului	Rămâne detaşat în abordarea problemei studiate, luând măsuri active pentru a-şi menţine obiectivitatea.	Se implică subiectiv în temă, uneori până la nivelul de a experimenta la rândul său trăirile asociate problemei studiate.
Focusul în cercetare	Focalizare pe descrierea şi explicarea comportamentului.	Focalizare pe înţelegerea şi interpretarea trăirilor asociate comportamentului.
Metodologie utilizată	Predominant cantitativă, cu sublinierea diferenţelor dintre fapte şi opinii.	Predominant calitativă, fără a diferenţia clar între fapte şi opinii.
Evaluarea validităţii concluziilor unei cercetări	Ceea ce contează este valoarea de adevăr a concluziilor obţinute.	Ceea ce contează este măsura în care insighturile şi interpretările noi oferite ajută la o mai bună înţelegere a fenomenului şi sunt împărtăşite de către participanţi.
Contribuţia la cunoaştere	Sunt apreciate îndeosebi cercetările care evidenţiază legităţi, adevăruri universale, în acest sens avantajate	Nu se urmăreşte oferirea unui răspuns clar la întrebarea cercetării, ci se preferă ca interpretarea fenomenului să conducă

	fiind cercetările fundamentale (de tip cauzal-explicativ).	la noi insighturi în legătură cu tema studiată.
Ilustrări ale unor teme de studiu posibile		
Satisfacţia profesorilor	Relaţia dintre creşterea numărului de elevi dintr-o clasă şi satisfacţia cadrelor didactice.	O investigaţie narativă a percepţiilor şi emoţiilor profesorilor ca urmare a modificării dimensiunii unei clase.
Tulburări de erecţie	Un studiu metaanalitic cu privire la eficienţa psihoterapiei în tratarea disfuncţiilor erectile.	Ce semnifică pentru bărbaţi şedinţele de psihoterapie pentru disfuncţiile erectile? O abordare fenomenologică.
Fluctuaţia de personal	Afectivitatea negativă ca predictor al fluctuaţiei de personal.	Investigarea secvenţelor emoţionale prin care trec angajaţii care părăsesc locul de muncă.

3.3.3. Cercetări cantitative vs cercetări calitative

Este posibil ca un număr important de cititori să rămână cu impresia că aspectele menţionate în paragrafele anterioare se potrivesc pentru a distinge şi între studiile cantitative şi cele calitative, pornind de la faptul că abordarea cantitativă este opţiunea standard pentru perspectiva pozitivistă, în timp ce metodologia calitativă este opţiunea de start în studiile bazate pe paradigma interpretativă. Cu toate acestea, lucrurile nu stau întotdeauna aşa. Diferenţierea dintre studii de tip cantitativ şi cele de tip calitativ are în vedere metodologia utilizată în colectarea datelor, în timp ce tipul de paradigmă abordată influenţează modul în care sunt analizate şi interpretate rezultatele. Astfel, este posibil ca un studiu care aplică o metodologie de cercetare de tip calitativ (interviuri aprofundate realizate pe câţiva participanţi-cheie) să apeleze la analiza de conţinut (metodă cantitativă) şi să interpreteze exclusiv rezultatele obţinute în

urma acestei analize (o manieră de lucru specifică pentru abordarea pozitivistă). Astfel, nu se poate pune un semn de egalitate între cercetarea calitativă şi abordările de tip interpretativ, pentru că datele calitative colectate pot fi integrate într-o manieră pozitivistă. Reversul este de asemenea teoretic posibil, deşi este mai puţin întâlnit în practică. De exemplu, într-o cercetare cu un obiectiv explorator poate fi aplicată o metodologie de tip cantitativ (aplicare de chestionare în cadrul unui sondaj efectuat pe un eşantion reprezentativ), însă interpretarea să fie una de tip hermeneutic. Se poate ajunge aici dacă: (a) accentul în studiu este pus mai degrabă pe interpretarea şi înţelegerea fenomenului studiat decât pe descrierea lui; (b) cercetătorul abordează într-o manieră subiectivă tema, interpretând rezultatele prin prisma opiniilor şi cunoştinţele sale anterioare despre subiect (mergând dincolo de ceea ce îi permit datele colectate); (c) îşi construieşte discursul (concluziile) într-o manieră mai degrabă hermeneutică, context în care se foloseşte de datele cantitative colectate ca temeiuri pentru comentariile sale de ordin calitativ.

Apoi, alegerea unei metodologii de lucru de tip cantitativ sau calitativ este dependentă şi de gradul de cunoaştere al subiectului abordat. Dacă tema investigată se referă la un subiect despre care se cunosc puţine lucruri, este preferabil un demers de tip explorator, situaţie în care metodologia de tip calitativ are un avantaj prin flexibilitatea ridicată cu privire la tipul de date colectate. În schimb, dacă subiectul abordat este unul cunoscut, demersul de cercetare depăşeşte stadiul unui studiu explorator, caz în care preferabilă este metodologia de tip cantitativ.

Prin urmare, în decizia de alegere a unei metodologii cantitative sau calitative intră doi factori importanţi: (a) gradul de cunoaştere a temei cercetate, decurgând din aceasta obiectivul de cercetat; (b) asumpţiile filosofice ale cercetătorului (aplecare spre a descrie şi explica realitatea

observabilă sau spre a interpreta şi acorda semnificaţii aspectelor [realităţii] construite).

O serie de cercetători precum Kazdin (2010), Elliot şi Timulak (2005) sau Bedos, Pluye, Loignon şi Levine (2009) au subliniat principalele diferenţe dintre metodologia cantitativă şi cea calitativă, cele mai importante fiind reflectate în tabelul 3.2. Aceste informaţii sunt utile pentru a oferi celor interesaţi o privire de ansamblu pentru a aprecia care metodologie ar trebui utilizată în funcţie de problema investigată (obiectivul studiului).

Tabelul 3.2. *Sumar al particularităţilor metodologiei cantitative* vs *calitative*

Aspect evaluat	Metodologia cantitativă	Metodologia calitativă
Obiectivul studiului	De obicei este unul inferenţial, bazat pe formularea şi testarea unor ipoteze de cercetare menite să identifice asocieri între variabile sau diferenţe între grupuri.	Este unul explorator, cu rolul de a descrie şi interpreta (sau oferi insighturi pentru) problema studiată, fără a testa ipoteze de cercetare.
Stadiul cunoaşterii problemei studiate	Aplicabilă îndeosebi asupra unor probleme cercetare anterior, despre care există deja o serie de cunoştinţe. Tema investigată este formulată adesea într-o manieră închisă, exactă (vezi modelul PICO amintit în primele două capitole).	Aplicabilă îndeosebi în studiul unor probleme noi, despre care nu se cunoaşte nimic sau se cunosc informaţii insuficiente. Tema investigată este formulată vag, adesea sub forma unei întrebări cu răspuns deschis (de exemplu, „În ce mod resimt adolescenţii plecarea părinţilor la muncă în străinătate?”).
Stilul de lucru în cercetare	Importanţă acordată controlului variabilelor cu potenţial confundat şi riscurilor rezultate din tipul de design utilizat, pentru a reduce numărul de explicaţii posibile (explicaţii	Variabilele relevante sunt studiate în contextul lor natural, fără a încerca să izoleze influenţa acestora de alţi factori cu potenţial de variabile confundate. Accentul se pune pe întreg, pe complexitatea descrierilor şi interpretărilor, pe sensuri multiple

	alternative pentru rezultatul obţinut).	care pot rezulta din datele colectate.
Numărul de participanţi	Număr mare pentru a asigura o putere statistică adecvată în procesul de testare a ipotezelor.	Număr mic de participanţi, compensat prin studierea lor în manieră intensivă (vezi tipul de date colectate).
Tipul datelor colectate	Date numerice, preferabil prin intermediul unor instrumente standardizate.	Descrieri narative, istorisiri, care nu sunt ulterior prelucrate în termeni cantitativi (analiza de conţinut fiind o metodă cantitativă).
Flexibilitate metodologică	Scăzută. De obicei întreaga procedură este una standardizată, algoritmică, mergând în cazul experimentelor până la măsuri cum ar fi contrabalansarea, pentru a contracara distorsionări potenţiale în rezultate.	Medie. Există o planificare iniţială, care poate fi ajustată pe parcursul colectării datelor, în funcţie de nevoi (de exemplu, în cazul utilizării interviurilor sunt preferate cele nestructurate sau semistructurate).
Analiza datelor colectate	Prin intermediul tehnicilor statistice, ţinând seama în interpretare de limitările impuse de designul de cercetare utilizat.	Analiză de text cu scopul: (a) de a descrie (de exemplu, a stabili teme centrale asociate problemei studiate din perspectiva trăirilor participanţilor); (b) de a interpreta/oferi sensuri şi semnificaţii acestor trăiri, prin apelul la filtrul subiectiv sau amprenta cercetătorului şi la teoriile naive ale participanţilor chestionaţi.
Aprecierea calităţii demersului de cercetare	Sunt preferabile studiile ale căror rezultate sunt replicate şi care conduc la teorii care să ne ajute să descriem şi explicăm mai bine aspectele psihologice investigate.	Sunt preferate studiile ce conţin descrieri care captează esenţa fenomenului studiat. Rezultatul este împărtăşit de ceilalţi şi poate fi replicat şi/sau obţinut din surse multiple (triangulare), accentul fiind pus pe înţelegerea aprofundată a fenomenului.

Desigur, aşa cum în cazul metodologiei cantitative există diferenţe mari de la un anumit tip de studiu la altul (de exemplu, între un sondaj de opinie şi un studiu experimental), în mod similar, sub umbrela metodologiei calitative se regăsesc o multitudine de abordări posibile. Creswell (2007) aprofundează cinci asemenea direcţii: (a) studiile

narative (accentul cade pe intervievarea subiecţilor pentru a relata istorioare din viaţa lor [istorii orale], urmate de realizarea unei sinteze literare în care să fie relatată viaţa indivizilor); (b) studiile fenomenologice (accentul este pus pe surprinderea esenţei experienţei studiate); (c) studiile etnografice (accentul se face pe interpretarea experienţelor respective prin prisma culturii împărtăşite de grupul din care indivizii fac parte); (d) studiile fundamentate (*grounded theory*) (în care se accentuează ideea generării unei teorii/perspective despre experienţa analizată pe baza opiniilor colectate de la participanţii la studiu) şi (e) studiile de caz de tip calitativ (în care predomină interesul de a obţine o analiză profundă a experienţei/individului, preferabil prin prisma unei teorii, făcând apel la surse multiple de documentare).

În sfârşit, merită subliniată expansiunea unei abordări ştiinţifice noi, cunoscută sub denumirea de metodologie mixtă (Tashakkori şi Teddlie, 1998), designul metodelor de cercetare mixte (Creswell, 2006) sau, mai frecvent, metode de cercetare mixte (*mixed method research*) (Creswell, 2006)[5]. Aceasta încearcă să combine demersul cantitativ, cu cel calitativ, pentru a beneficia de avantajele strategice ale ambelor metodologii (o înţelegere aprofundată a constructului studiat – abordarea calitativă; controlul variabilelor străine şi o creştere a nivelului de generalizare a rezultatelor – abordarea cantitativă) (Creswell, 2006; Steckler *et al.*, 1992).

Figura 3.1 prezintă patru situaţii în care este recomandabilă utilizarea unui design mixt, primele două în care metodologia calitativă este utilizată ca un adjuvant la cea cantitativă, cea de-a treia referindu-se la reversul situaţiei, iar ultima, ca un posibil exemplu de triangulare, astfel încât rezultatul obţinut să fie sprijinit prin utilizarea ambelor tipuri de metodologie.

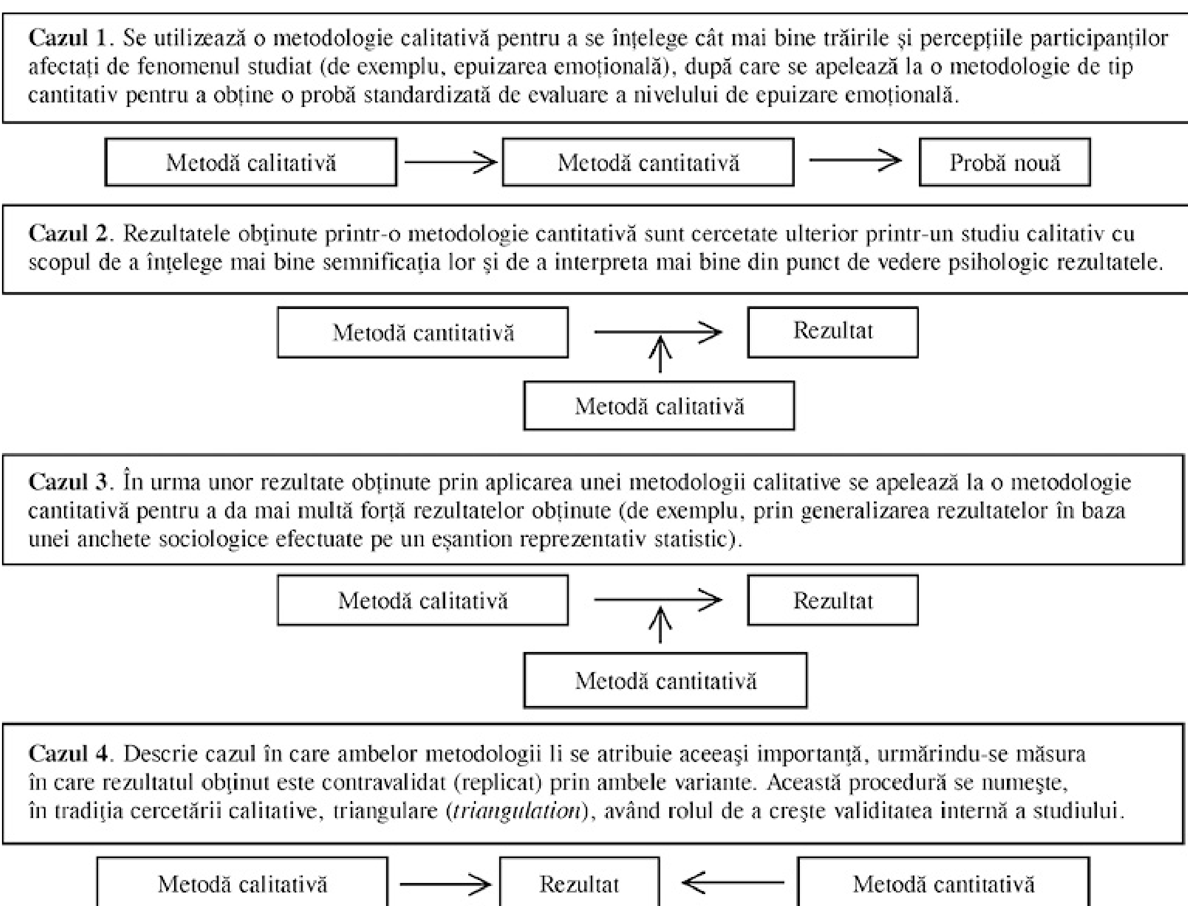


Figura 3.1. *Ilustrarea celor mai întâlnite tipuri de design mixt. Adaptare după Steckler* et al. *(1992)*

3.3.4. Cercetări experimentale vs cercetări cvasiexperimentale vs cercetări non-experimentale

Ultimele două clasificări ale tipurilor de cercetări menţionate la 3.3.2 şi 3.3.3 au drept caracteristică comună faptul că se bazează pe diferenţele existente la cel mai general nivel, cel al metodologiei cercetării. În practică se întâmplă rar să se ia decizii la acest nivel de generalitate, de aceea probabil că 99% dintre cercetările psihologice existente îmbrăţişează paradigma dominantă, aceea a unui metodologii pozitiviste

de tip cantitativ. În acord cu această stare de fapt a fost proiectată structura acestei cărţi, accentul fiind pus pe metodologia cantitativă de orientare pozitivistă, pentru a putea răspunde fluxului principal de preocupări din domeniul cercetării psihologice.

În schimb, clasificarea tipurilor de studii în cercetări experimentale *vs* cvasiexperimentale *vs* non-experimentale, vizează cel de-al doilea nivel de generalitate, cel al tipului de design al cercetării. Astfel că în acest caz, spre deosebire de cazul celorlalte clasificări menţionate, cercetătorul joacă un rol activ în determinarea tipului de cercetare efectuată. Acest lucru se întâmplă pentru că la baza clasificării studiilor în cele trei categorii se are în vedere nivelul de control exercitat de către cercetător asupra variabilelor străine. Cu cât se exercită un control mai drastic într-un studiu asupra acestor variabile potenţial confundate, cu atât creşte încrederea în concluziile cercetării respective.

În termeni tehnici există două măsuri cu rol de control, cu caracter obligatoriu, care trebuie luate din perspectiva designului unei cercetări pentru a creşte încrederea în validitatea rezultatelor studiului. Acestea privesc exercitarea unui control riguros asupra variabilelor potenţial confundate prin: (a) manipularea variabilei independente cu scopul de a varia efectul acesteia asupra variabilei dependente (de exemplu, se evaluează două forme diferite de intervenţie şcolară menite să diminueze abandonul şcolar); (b) distribuirea aleatorie a participanţilor la studiu în cele două tipuri de intervenţie, cu scopul de a egaliza (controla) efectul pe care alte variabile potenţial confundate îl poate avea asupra variabilei dependente. Dacă sunt îndeplinite amândouă cerinţele, vorbim despre studii experimentale. Dacă niciuna dintre condiţii nu este îndeplinită vorbim despre studii non-experimentale. În sfârşit, dacă doar condiţia (a) este îndeplinită vorbim despre studii cvasiexperimentale. Acestea mai

sunt cunoscute şi sub denumirea de studii experimentale „slabe”, fiind încadrate de anumiţi autori în categoria generală a experimentelor, ca urmare a rolului activ pe care îl joacă cercetătorul prin manipularea variabilei independente (Shadish *et al.*, 2002).

Detalii suplimentare despre studiile experimentale (inclusiv cele cvasiexperimentale), respectiv despre studiile non-experimentale sunt oferite pe larg în capitole distincte ale acestei cărţi (capitolele 5 şi 6). Am rezervat acest spaţiu larg de prezentare a caracteristicilor acestor tipuri de studii datorită impactului ridicat pe care îl au asemenea informaţii în dezvoltarea gândirii critice cu referire la calitatea unei cercetări şi în planificarea studiilor empirice din psihologie. Aceste competenţe tehnice sunt esenţiale în realizarea de cercetări, fiind fundamentale pentru planificarea şi derularea în bune condiţii a unui studiu.

3.3.5. Cercetări exploratorii vs cercetări descriptive vs cercetări cauzale

Distincţia dintre studii experimentale şi cele non-experimentale este prezentă în orice manual introductiv de psihologie experimentală/metodologia cercetării. În schimb, mai puţin cunoscută este o altă taxonomie importantă pentru dezvoltarea gândirii critice legate de cercetări, ce are în vedere clasificarea lor în funcţie de tipul de obiectiv vizat de studiu (scopul studiului).

Clasificarea cercetărilor în funcţie de tipul de obiectiv vizat de studiu este esenţială, deoarece, aşa cum remarca Brewer (2000), evaluarea validităţii unui studiu, adică a nivelului de adevăr al afirmaţiilor rezultate în baza studiului, nu se face *per se*, ci în funcţie de tipul de obiectiv urmărit. De aceea, cu cât obiectivul este unul mai ambiţios (de exemplu,

să explice de ce apare un fenomen, nu doar să-l descrie), cu atât sunt necesare mai multe măsuri de control la nivelul designului studiului. Prin urmare, tipul de obiectiv studiat are impact direct asupra designului ales (de exemplu, studiul non-experimental de tip corelaţional etc.). Tabelul 3.3. oferă o descriere succintă a tipurilor de cercetări în funcţie de obiectivul studiului, iar figura 3.2 oferă o privire de ansamblu asupra legăturii dintre cele două clasificări avute în vedere (3.3.4 şi 3.3.5), luând în considerare alte consecinţe majore implicate la nivel de design (evaluate la un nivel inferior [mai concret] de analiză decât cel menţionat în secţiunea 3.3.4).

Tabelul 3.3. *Tipuri de studii din perspectiva obiectivului cercetării. O privire sintetică*

Clasificare după obiectivul vizat	Descriere succintă (tipuri de întrebări ale cercetării vizate)
1. Exploratoriu	Se referă la studii bazate pe metodologii calitative (mai frecvent) sau pe metodologii cantitative asupra unei teme despre care nu se cunoaşte nimic sau se cunosc foarte puţine lucruri. Două atribute specifice ale acestor studii sunt: (a) studiul nu are prestabilite ipoteze care să fie testate (de aici caracterul explorator); (b) deşi are libertatea de a-şi propune, fie să descrie, fie să prezică, fie chiar să ofere explicaţii pentru fenomenele studiate, concluziile din aceste studii sunt provizorii (nu permit stabilirea cu claritate a valorii de adevăr a concluziilor studiului). Aceste concluzii provizorii pot deveni ipoteze de cercetare ce urmează a fi testate în studii ulterioare.
2. Descriptiv 2A. Descriptiv propriu-zis 2B. Analitic	Se referă la studii bazate preponderent pe metodologii de tip cantitativ, non-experimentale (observaţionale), realizate cu scopul de *a descrie* cât mai acurat fenomenul studiat prin cercetări de teren („care este prevalenţa cazurilor de ADHD printre adolescenţii cu părinţi plecaţi la muncă în străinătate?” (2A), „câte subtipuri de ADHD există?” (2A), „care sunt modalităţile adaptive, respectiv dezadaptive de răspuns (coping) în situaţiile de epuizare emoţională?” (2A), „există o legătură între nivelul de agresivitate şi cogniţiile iraţionale autodepreciative?” (2B), „nivelul stimei de sine al persoanelor care au suferit un eşec şcolar este mai redus decât cel din lotul celor fără eşec şcolar etc.” (2B). În cazul subtipului „studii descriptive propriu-zise” (2A), analizele statistice efectuate se rezumă la indicatori descriptivi, nefiind formulate ipoteze de

	cercetare. Printre cele mai întâlnite studii de acest gen sunt studiile de caz[6], sondajele de opinie şi studiile bazate pe metoda observaţiei. În cazul subtipului de obiective descriptive analitice (2B), studiile efectuate conţin ipoteze de cercetare verificate prin intermediul tehnicilor statistice inferenţiale. Cel mai frecvent, pentru a răspunde la aceste obiective se apelează la studiile non-experimentale de tip corelaţional sau la cele de tip comparativ, non-experimental (*case control studies*). Pornind de la *descrierea* asocierilor observate, studiile descriptiv-analitice permit formularea de *predicţii*, de exemplu, cu privire la relaţia dintre variabila studiată şi alte variabile de interes (de exemplu, identificarea factorilor de risc care contribuie la apariţia anxietăţii sociale). Identificarea predictorilor nu este însă echivalentă cu identificarea cauzei pentru apariţia fenomenului avut sub observaţie, deoarece predictorul poate juca orice rol în asocierea cu criteriul de interes, inclusiv cel de variabilă confundată. De exemplu, dacă nivelul de studii (numărul de ani de şcoală) prezice preferinţa pentru un anumit post muzical de radio, ca urmare a asocierii dintre cele două variabile, nu înseamnă însă că educaţia reprezintă şi sursa (cauza) acestei preferinţe.
3. Cauzal 3A. Cauzal-descriptiv 3B. Cauzal-explicativ	Se referă, în exclusivitate, la studii bazate pe o metodologie cantitativă riguroasă, (experimente şi, eventual, cvasiexperimente) interesate de identificarea unor *relaţii cauzale* şi a unor *explicaţii* (mecanisme cauzale) responsabile pentru apariţia efectului urmărit. Prin acest tip de obiective se trece dincolo de simpla descriere a unui comportament (fenomen), accentul fiind pus pe oferirea de explicaţii pentru dinamica comportamentului studiat. În cercetarea aplicată, axată pe aspecte de utilitate, un scop de tip cauzal-descriptiv este suficient. Cu ajutorul experimentelor se vizează doar demonstrarea unei relaţii de tip cauză-efect („coaching-ul cognitiv-comportamental este eficient în reducerea stresului ocupaţional”), fără a valida ştiinţific explicaţia pentru care acest lucru se întâmplă. În cercetarea fundamentală, axată pe a găsi explicaţii (de ce este coaching-ul cognitiv-comportamental eficient în diminuarea stresului ocupaţional?), se apelează la scopuri cauzal-explicative, care urmăresc să expliciteze mecanismul cauzal. În acest caz, accentul este pus pe demonstrarea relaţiei cauzale, dar mai ales pe explicarea (identificarea) ingredientului activ („de ce şi în ce context funcţionează relaţia cauzală?”).

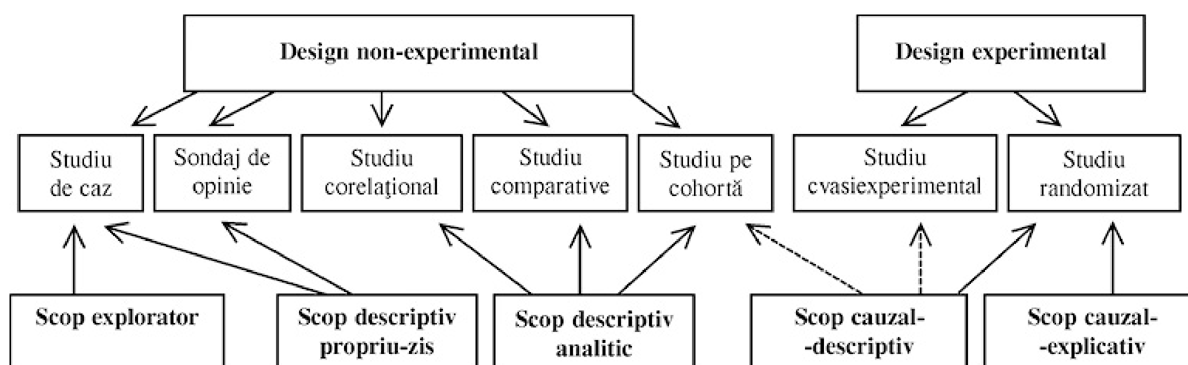


Figura 3.2. *Implicaţii în alegerea tipului de studiu în funcţie de designul cercetării şi de scopul urmărit*

Se observă, ca aspect de principiu, că obiectivul (scopul) unui studiu poate fi atins numai prin anumite tipuri de studii, în funcţie de particularităţile designului pe care îl implică acesta. Bunăoară, în cazul studiilor cu obiective de tip descriptiv sunt disponibile mai multe alternative de studiu, din categoria celor care prezintă un design non-experimental. De exemplu, pentru a *descrie relaţia* dintre două variabile, cum ar fi dacă există o asociere între satisfacţia în relaţia de cuplu şi nivelul de romantism, este suficient să apelăm la un studiu corelaţional, ca model potrivit pentru scopuri de tip descriptiv-analitic. Nu este nevoie să apelăm la designuri mai complicate şi mai costisitoare, de tipul studiilor cu design experimental, pentru că se poate răspunde la întrebare şi prin studii mai uşor de implementat, cum sunt cele cu un design non-experimental. De asemenea, este mai puţin util să se apeleze la studii care urmăresc obiective exploratorii, deoarece concluziile obţinute prin aceste studii au un caracter provizoriu, generând ipoteze care ar trebui validate ulterior prin alte studii cu scop descriptiv.

În schimb, dacă ceea ce se urmăreşte prin studiu este de a identifica *o posibilă relaţie cauzală*, studiul corelaţional se dovedeşte a fi insuficient pentru că nu poate tranşa clar sursa corelaţiei observate (care dintre următoarele variante este mai apropiată de adevăr): (a) în relaţiile de

cuplu cu un nivel ridicat de satisfacţie apar mai frecvent dovezi de romantism (satisfacţia maritală cauzează intensificarea romantismului); (b) prin creşterea frecvenţei actelor romantice sporeşte satisfacţia maritală (romantismul conduce la o satisfacţie maritală mai mare); (c) între cele două variabile există o intercondiţionare reciprocă, asemenea unui cerc „virtuos" (pe măsură ce sporesc comportamentele romantice creşte satisfacţia maritală, iar aceasta induce, la rândul ei, o predispoziţie spre şi mai multe comportamente romantice etc.); (d) ambele variabile studiate sunt efecte ale duratei relaţiei de cuplu (cuplurile formate mai recent percep un nivel mai ridicat de romantism şi un nivel mai ridicat al satisfacţiei în relaţia de cuplu).

Pentru a elimina dintre variantele menţionate pe acelea care nu sunt sprijinite de date, va trebui să adoptăm un alt tip de studiu decât cel corelaţional transversal. De pildă, pentru a observa gradul de sprijin existent pentru cele menţionate la punctul (d), ar fi util un studiu longitudinal care ar presupune urmărirea pe o anumită perioadă de timp a modului în care variază cele două variabile, pe măsură ce durata relaţiei de cuplu creşte. Însă cea mai bună variantă de a verifica dacă există o relaţie de tip cauzal este apelul la studii experimentale, dacă acestea sunt permise din punct de vedere etic. În exemplul oferit ar putea fi verificată astfel doar varianta (b), deoarece nivelul comportamentului romantic poate fi manipulat experimental, în schimb ar fi inadecvat din punct de vedere etic să manipulăm nivelul satisfacţiei în relaţia de cuplu. Spre exemplu, participanţii la un asemenea studiu ar putea fi distribuiţi randomizat în două grupe experimentale: una care să cuprindă persoane care sunt instruite de către un consilier de cuplu cum anume să-şi surprindă partenerul prin acte romantice, ceilalţi incluşi într-un grup de tip placebo (discuţii purtate cu consilierii de cuplu despre problemele lor

maritale, cu scopul de a facilita catarsisul; consilierii rezumându-se la un rol suportiv, asemenea unui prieten). Dacă la finalul intervenţiei primul grup (cel care a deprins şi exersat abilitatea de a se comporta în mod romantic faţă de partener) va avea un nivel mai ridicat de satisfacţie în relaţia de cuplu decât cel de-al doilea grup, există dovezi empirice cu privire la rolul cauzal pe care îl joacă comportamentul romantic în creşterea satisfacţiei faţă de relaţia de cuplu. Atunci când apelul la studii experimentale nu este posibil (din motive etice sau logistice), se pot apela la variante de suplinire cum sunt studiile cvasiexperimentale sau studiile longitudinale (vezi liniile întrerupte din figura 3.2). Acestea pot oferi indicii în sprijinul sau împotriva unei relaţii de tip cauzal, însă nu sunt la fel de sigure precum studiile experimentale în identificarea unei relaţii cauzale.

În cazul în care o asemenea relaţie cauzală este sprijinită empiric, cercetările ulterioare îşi pot propune obiective de tip explicativ pentru a identifica mecanismele cauzale responsabile pentru efectul observat. În exemplul nostru ipotetic, dacă se demonstrează experimental că comportamentul romantic determină o creştere a satisfacţiei maritale, se pot concepe studii de tip experimental care să ofere explicaţii valide pentru această relaţie cauzală. Orice alt tip de studiu, în afara celui experimental, nu beneficiază de suficientă rigurozitate metodologică pentru a elimina probleme specifice variabilelor cu potenţial confundat. Spre exemplu, printre explicaţiile posibile pentru relaţia cauzală constatată putem identifica: (a) comportamentele romantice conduc la intensificarea percepţiei că partenerul ne iubeşte sincer, iar acest lucru conduce la o creştere a satisfacţiei maritale (afectivitatea recepţionată ca factor explicativ pentru creşterea satisfacţiei maritale în urma comportamentelor romantice); (b) relaţia cauzală este determinată de

activarea normei de reciprocitate în cazul partenerului (comportamentul nostru romantic conducând la reacţii pozitive din partea partenerului, care la rândul lui răspunde prin gesturi pozitive, fapt care conduce la o intensificare a satisfacţiei cu privire la relaţia de cuplu – norma de reciprocitate ca factor explicativ pentru creşterea satisfacţiei maritale în urma adoptării unui comportament romantic). Seria de posibile explicaţii alternative pentru relaţia cauzală identificată ar putea continua. Ceea ce ne interesează aici este sublinierea faptului că singura modalitate de a trece de la nivelul speculativ de enumerare a unor posibile explicaţii pentru relaţia cauzală identificată la stadiul de a valida ştiinţific (de a stabili valoarea de adevăr a acestor explicaţii) este apelul la studii experimentale, în care să fie manipulate, pe rând, explicaţiile propuse. Ideal, ar putea fi propuse studii în care să fie testate (comparate direct) două sau mai multe asemenea explicaţii posibile, însă dacă acest lucru nu este posibil, explicaţiile respective pot fi evaluate succesiv, printr-o serie de studii experimentale. Nu vom insista aici asupra dezvoltării exemplului menţionat în cazul unor studii cu obiective explicative, deoarece cititorii interesaţi pot găsi o asemenea ilustrare în caseta 4.2 (din capitolul 4).

În schimb, vom preciza două caracteristici ale demersurilor de tip cauzal-explicativ, menţionate de Shadish *et al.* (2002): tendinţa spre abordare moleculară în locul uneia molare, respectiv tendinţa spre abordarea multinivelară. Prima se referă la tendinţa întâlnită de a descompune factorul-cauză în subcomponente pentru a diferenţia care anume dintre ingredientele respective reprezintă factorul activ. Dacă în cercetarea cu scop cauzal-descriptiv abordarea molară este suficientă (de exemplu, de a arăta că terapia raţional emotivă este eficientă în combaterea depresiei), în cercetarea în scop cauzal-explicativ se caută a se evidenţia ce anume din ceea ce se realizează în cadrul terapiei raţional-

emotive reprezintă mecanismul schimbării, fiind responsabil pentru ameliorarea simptomatologiei depresive („este vorba doar de un factor comun, cum ar fi încrederea în psihoterapeut?” sau „sunt anumite elemente specifice cum ar fi modificarea cogniţiilor iraţionale autodepreciative?” etc.). Identificarea elementului activ sau a setului de elemente active este un aspect esenţial, nu numai pentru procesul cunoaşterii în sine (facilitează dezvoltarea de noi teorii cu privire la modul de funcţionare psihică), ci şi pentru a putea proiecta intervenţii optimizate (care să valorifice mai bine ingredientele active responsabile pentru efectul dorit, cum ar fi o variantă de psihoterapie raţional-emotivă care să accentueze exerciţiile de disputare şi modificare a cogniţiilor iraţionale autodepreciative în unele raţionale, dacă se dovedeşte că aceste cogniţii stau la baza concluziilor din studiile experimentale focalizate spre a arăta de ce psihoterapia raţional-emotivă este eficientă în combaterea depresiei.

Cea de-a doua caracteristică specifică abordării cauzal-explicative vizează abordarea multinivelară. Astfel, mecanismele cauzale analizate pot viza niveluri diferite de generalitate, respectiv de focalizare pe aspecte implementaţionale (biologice). În exemplul amintit pot fi căutate explicaţii de sorginte neuropsihologică (modificări la nivelul unor neurotransmiţători chimici), de natură psihologică abstractă (nivelul general de iraţionalitate cognitivă) sau de natură psihologică concretă (cogniţiile iraţionale de tip autodepreciative). De asemenea, explicaţiile pot viza diferite cadre conceptuale (raportarea la diferite teorii), în funcţie de teoria psihologică îmbrăţişată de cercetător, putând fi oferite explicaţii de sorginte psihodinamică, a terapiei familial-sistemice etc. Acestea pot fi considerate valide, atâta timp cât sunt sprijinite empiric de studii care utilizează un design adecvat (în speţa de faţă, unul experimental) pentru a

răspunde obiectivului studiului, acela de a oferi explicaţii valide ştiinţific pentru relaţia cauzală observată.

Un alt aspect care ar trebui subliniat cu privire la figura 3.2 vizează pleiada de studii care fac apel la un design non-experimental. Numărul de studii care apelează la un design non-experimental este mai mare decât cel prezentat în figura respectivă (de exemplu, lipsesc studiile care apelează la o metodologie calitativă). Capitolul 5 este destinat descrierii acestei categorii de studii denumite generic studii non-experimentale. Ceea ce ar trebui subliniat aici este faptul că studiile bazate pe un design non-experimental pot fi utilizate interschimbabil, în genere, atât pentru obiective de tip descriptiv, cât şi pentru cele de tip explorativ. De exemplu, studiul de caz, aşa cum remarca şi Kazdin (2010), se poate dovedi util atât în scopuri exploratorii (ca sursă de idei şi ipoteze pentru studii viitoare), cât şi în scopuri descriptive propriu-zis (pentru a ilustra, în scop educaţional sau de *advocacy*, cazuri particulare în sprijinul unei teorii sau unei intervenţii psihologice; pentru a descrie aspecte psihologice rar întâlnite).

De asemenea, un tip aparte de studii non-experimentale (studiile longitudinale dintre care cel mai întâlnit tip sunt studiile pe cohorte) poate fi utilizat pentru a infera afirmaţii de tip cauzal, atunci când datele experimentale lipsesc sau nu pot fi colectate din motive etice sau logistice. Acest lucru este posibil pentru că una dintre premisele unei relaţii de tip cauză-efect este aceea de a observa că modificările la nivelul variabilei-cauză precedă temporal modificările la nivelul variabilei-efect, iar observarea acestui lucru este posibilă doar într-un design longitudinal, ce presupune observarea evoluţiei participanţilor la studiu de-a lungul timpului.

La finalul acestei secţiuni merită să subliniem caracterul cumulativ al cunoaşterii, cu implicaţii asupra tipului de studii care ar trebui utilizate în funcţie de stadiul cunoaşterii în cazul temei studiate. Astfel, studiile care urmăresc un obiectiv cauzal-explicativ nu îşi au sensul până când nu există dovezi ale existenţei unei relaţii cauzale, abordabilă prin obiective de tip cauzal-descriptiv. De asemenea, nu are rost să căutăm a identifica relaţii de tip cauză-efect ca atare (obiectiv descriptiv) în absenţa unor dovezi cu privire la existenţa unei relaţii. În sfârşit, în multe cazuri, înainte de a avea studii descriptive cu un demers coerent (în care să ştim ce să căutăm) ar fi util să apelăm la studii exploratorii pentru a ne face o impresie despre ce anume să studiem ulterior într-o manieră sistematică. Cu alte cuvinte, deşi experimentele sunt cele mai riguroase din punct de vedere metodologic, ele sunt cele mai dificil de implementat şi nu îşi au sensul în absenţa unor dovezi empirice deja existente cu privire la descrierea fenomenului studiat. În consecinţă, studiile experimentale în psihologie sunt mai potrivite pentru a adresa întrebări de cercetare mature, în timp ce studiile de tip non-experimental sunt utile pentru a strânge, mai degrabă într-o manieră nesistematică, acea cantitate de informaţii necesară pentru a permite saltul în cunoaşterea domeniului atât prin evaluări sistematice sau narative ale literaturii de specialitate din domeniu, cât şi prin trecerea la nivelul următor de interes al oricărui demers ştiinţific, cum ar fi trecerea de la descriere către predicţie şi către oferirea de explicaţii pentru problema studiată.

3.4. Aprecierea validităţii afirmaţiilor din studiile efectuate

În esenţă apelăm la ştiinţă ca sursă a cunoaşterii în defavoarea oricărei alte surse de cunoaştere (apelul la autoritatea academică, la convingeri personale etc.) pentru că ne aşteptăm ca prin apelul la demersul ştiinţific să ajungem să cunoaştem realitatea (răspunsul corect la întrebarea de cercetare adresată) şi să formulăm concluzii solide, fundamentate despre problematica studiată.

Validitatea unei cercetări se poate referi, în cazuri rare, la studiul în ansamblul său (dacă designul utilizat este total inadecvat pentru a răspunde la întrebarea adresată prin cercetare), însă cel mai frecvent are în vedere orice afirmaţie extrasă în baza cercetării realizate, apreciind măsura în care studiul respectiv, prin caracteristicile sale metodologice, poate acoperi valoarea de adevăr a afirmaţiei în cauză (se poate stabili fără doar şi poate că afirmaţia respectivă este una adevărată, validă). De aceea, aşa cum afirma Brewer (2000), discuţia despre validitate nu se referă la cercetarea în sine, ci priveşte mai degrabă validitatea studiului în raport cu obiectivul urmărit prin studiu (vezi caseta 3.1).

Caseta 3.1. *Ilustrări ale impactului pe care obiectivul urmărit (şi implicit tipul de studiu) îl are asupra validităţii afirmaţiilor rezultate din studiu*

Exemplul 1. Identificarea măsurii în care există o legătură între nivelul de violenţă din desenele animate urmărite şi nivelul de agresivitate al copiilor (context: studiu non-experimental transversal de tip corelaţional implementat corect, cu rezultate ce indică o asociere semnificativă statistic). Obiectiv vizat: descriptiv-analitic. Rezultat: nivelul de agresivitate al copiilor corelează pozitiv cu gradul de violenţă din conţinutul desenelor urmărite. **Concluzie (afirmaţie) validă** în baza observaţiei: fiind un studiu corelaţional, nu se poate stabili clar sursa acestei asocieri. Astfel este posibil fie ca desenele cu un conţinut violent să inducă o creştere a nivelului de agresivitate în rândul copiilor, fie să asistăm la un efect al autoselecţiei, prin care copiii cu un nivel mai ridicat de agresivitate să prefere

să urmărească mai frecvent desene cu conţinut violent, deoarece acestea le răspund mai bine nevoilor.
Concluzie (afirmaţie) nevalidată în baza observaţiei (studiul efectuat nu poate stabili dacă este corectă sau nu afirmaţia respectivă): acest rezultat sprijină poziţia grupurilor de *advocacy* de a condamna conţinutul violent din desene animate, observându-se că nivelul violenţei creşte pe măsură ce sunt vizionate mai multe desene cu conţinut violent. Explicaţia pentru această stare de fapt este aceea că copiii învaţă că actele violente pot constitui un mod uzual de comportament soldat cu succes (agresivitatea ca mijloc instrumental în atingerea scopului dorit).

Exemplul 2. Demonstrarea unei relaţii cauzale, conform căreia desenele animate cu un conţinut violent determină creşterea nivelului de agresivitate la copii (context: studiu experimental implementat corect, cu rezultate ce indică diferenţe semnificative între cele două grupe, în direcţia aşteptată).

Obiectiv vizat: cauzal-descriptiv.
Rezultat: un nivel mai ridicat al agresivităţii măsurate posttest în grupul experimental (cei care au vizionat desene cu conţinut violent) decât în grupul de control (cei care au vizionat desene fără conţinut violent).

Concluzie (afirmaţie) validată[*/**]: datele obţinute sprijină ideea că desenele cu un conţinut violent induc agresivitate în rândul copiilor. Acest rezultat sprijină poziţia grupurilor de *advocacy* de a condamna conţinutul violent din desene animate.
Concluzie (afirmaţie) nevalidată: explicaţia pentru această stare de fapt este aceea că copiii învaţă că actele violente pot constitui un mod uzual de comportament soldat cu succes (agresivitatea ca mijloc instrumental în atingerea scopului dorit).

* Am semnalat prin asterisc cazul respectiv pentru că doar apelul la un design potrivit (de exemplu, un studiu experimental) nu este suficient pentru a valida afirmaţii cauzale. Este la fel de important ca implementarea să fie una corectă. De pildă, dacă studiul este efectuat pe băieţi, iar grupul experimental urmăreşte desene din seria *Ben 10*, în timp ce copiii din grupul de control vizionează *Bambi*, diferenţele constate pot fi determinate de alte variabile confundate (cum ar fi nivelul de atractivitate, respectiv de familiaritate diferit pentru cele două desene, mai ridicat pentru *Ben 10* în cazul băieţilor) sau de gradul diferit de similaritate dintre jucăriile avute la dispoziţie şi personajele din desene (variabila dependentă fiind măsurată prin observarea comportamentului copilului prin joc, ulterior vizionării desenului). Astfel, dacă în grădiniţă sunt mai degrabă jucării din seria *Ben 10* sau alte personaje extraterestre similare, dar nu există căprioare sau alte personaje incluse în *Bambi*, există riscul ca efectul observat să fie ca urmare a facilitării

ridicate a transferului unor scenarii din desene în jocul de rol al copiilor în favoarea celor pentru care există recuzita necesară. Cu alte cuvinte, este important ca manipularea experimentală să fie una adecvată pentru a surprinde factorul cauzal anticipat şi pentru a controla alte variabile potenţial confundate (de exemplu, ar fi potrivit a se apela la două episoade din acelaşi serial de desene [aceleaşi personaje de bază], unul cu conţinut violent, altul fără violenţă).
** Cu toate măsurile de control ce pot fi luate pentru a elimina o serie de riscuri la adresa validităţii unui studiu, este imposibil să fie acoperite toate situaţiile posibile, motiv pentru care: (a) niciun studiu nu poate atinge perfecţiunea (lipsa totală a oricărui risc privitor la validitatea concluziilor sale); de exemplu, apelul la studii randomizate controlate, deşi necesar pentru a urmări un obiectiv de tip cauzal-descriptiv sau cauzal-explicativ, nu constituie o garanţie pentru validitatea afirmaţiilor ce fac referiri la aspecte cauzal-descriptive sau cauzal-explicative; (b) ideea de valoare de adevăr (rezultat valid) trebuie privită în termeni probabilistici (este foarte probabil ca afirmaţia să fie una validă, adevărată), nu în termeni absoluţi.

De aceea, în redactarea acestei secţiuni, pe care o consider fundamentală pentru a înţelege esenţa demersului de cercetare şi a permite evaluarea contribuţiei aduse de diferite studii, am apelat atât la surse clasice, cât şi la surse complementare. Dintre sursele clasice care tratează problematica validităţii se disting contribuţiile lui Shadish *et al.* (2002) sau Cook şi Campbell (1979), în special în descrierea principalelor riscuri privitoare la validitatea concluziilor unui studiu experimental sau cvasiexperimental. Însă, dincolo de această discuţie generală despre validitatea unei cercetări, vom introduce în discuţie surse complementare precum Kazdin (2010), al cărui principal merit este de a prezenta riscuri mai puţin cunoscute privitoare la validitatea concluziilor unui studiu, dar importante din perspectiva psihologiei validate ştiinţific (fundamentate empiric), cum sunt efectul de noutate sau planificarea asincronă a evaluărilor, atunci când se apreciază validitatea eficienţei unei intervenţii.

Mai mult, pornind de la afirmaţia repetată cu mai multe ocazii pe parcursul acestei cărţi, aceea că nu este posibil să fie realizată o cercetare

„perfectă”, lipsită de riscuri privitoare la validitatea concluziilor sale, voi aprofunda discuţia despre validitatea unui cercetări în funcţie de obiectivul studiului respectiv. Tipurile de validitate discutate au importanţă diferită în funcţie de scopul studiului. De exemplu, pentru un studiu experimental ce îşi propune un obiectiv cauzal-descriptiv, esenţial este ca acesta să răspundă îndeosebi aspectelor de validitate internă (demonstrarea relaţiei cauză-efect) şi secundar aspectelor care ţin de validitatea externă (nivelul de generalizare a relaţiei cauzale observate în alte contexte). În schimb, într-un studiu non-experimental ce îşi propune un obiectiv descriptiv, de tipul sondajului de opinie, se acordă o importanţă mai mare aspectelor ce ţin de validitatea de construct, respectiv cea externă, preocuparea fiind foarte redusă pentru problematica validităţii interne, pentru că prin definiţie aceasta ar urmări nivelul de adevăr al inferenţelor cu referire la relaţii de tip cauzal, aspect care nu poate fi surprins adecvat prin intermediul unui sondaj de opinie. În consecinţă, dincolo de prezentarea principalelor tipuri de validitate, ne propunem în secţiunea de faţă să extindem viziunea prezentată de Brewer (2000), cu privire la impactul pe care scopul urmărit de un studiu îl are asupra evaluării cercetărilor din perspectiva nivelului lor de validitate. Aceste informaţii vor fi utile atât în proiectarea cercetărilor viitoare (anticipând posibile probleme cu interpretarea rezultatelor în funcţie de caracteristicile studiului efectuat), cât şi în identificarea punctelor tari şi a punctelor slabe pentru studiile deja efectuate.

Tabelul 3.4. *Tipuri de validitate (adaptare după Shadish* et al.*, 2002)*

Tipuri de validitate	Descriere succintă (tipuri de afirmaţii vizate)
Validitate statistică	Se referă la: (a) validitatea inferenţelor statistice efectuate (măsura în care au fost evitate erori de tip I sau de tip II în testarea ipotezelor) şi la (b) încrederea în magnitudinea efectului observat.

Validitate internă	Se referă la validitatea inferenţelor conform cărora relaţia observată între A şi B reflectă situaţia unei relaţii cauzale (în ce măsură putem fi siguri că A îl determină pe B şi nu sunt alte variabile confundate responsabile pentru efectul B).
Validitate de construct	Se referă la măsura în care caracteristicile de bază ale studiului (variabile, participanţi etc.) reprezintă operaţionalizări potrivite pentru a surprinde constructele teoretice vizate.
Validitate externă	Se referă la măsura în care rezultatele obţinute pot fi generalizate pe alţi participanţi, alt context de cercetare sau alte instrumente de cercetare ce se referă la aceleaşi constructe. Cu alte cuvinte se referă la gradul de generalizare a rezultatelor (aceleaşi rezultate ar fi obţinute în alte studii similare).

3.4.1. Validitatea statistică

În limbaj simplu, această formă de validitate este valabilă pentru orice tip de studiu care utilizează o metodologie cantitativă şi îndeosebi pentru cele care apelează la formularea şi testarea unor ipoteze de cercetare (de exemplu, studii având un scop descriptiv-analitic). Mai concret, validitatea statistică are în vedere valoarea de adevăr a inferenţelor legate de: (a) măsura în care rezultatul semnificativ sau nesemnificativ statistic este unul corect (evitarea erorilor de tip I sau II, aspect dependent de puterea statistică a studiului); (b) măsura în care mărimea efectului observat în studiu este corect apreciată (efectul nu este supra- sau subestimat, respectiv încrederea în valoarea obţinută).

Dintre multiplele riscuri care pot afecta validitatea statistică se distinge îndeosebi situaţia unei puteri statistice scăzute de testare a ipotezelor. Puterea statistică scăzută reduce probabilitatea de a obţine rezultate semnificative statistic, când în realitate ar exista un asemenea efect, situaţie cunoscută sub numele de eroare de tip II. Puterea statistică scăzută poate fi cauzată de o multitudine de factori (vezi Sava

2004/2011), dintre care cel mai cunoscut este numărul insuficient de participanţi cuprinşi în studiu. Pentru a creşte puterea statistică a studiului, pe lângă creşterea volumului eşantionului studiat, se poate apela la măsuri de ordin metodologic precum: (a) utilizarea unui design intragrup în locul unuia de tip intergrup; (b) preferinţa, în cazul studiilor comparative, pentru designuri paralele cu mai mult de două grupe; (c) apelul la studii randomizate stratificate (*randomized blocked design*) în locul studiilor randomizate complet (*completely randomized design*) pentru a reduce eterogenitatea participanţilor comparaţi, cu condiţia ca variabila utilizată pentru stratificare/potrivire să fie una care corelează cu variabila dependentă; (d) utilizarea unor designuri cu număr egal de subiecţi în celulele experimentale etc. (Shadish *et al.*, 2002; Sava, 2004/2011).

Pe lângă măsurile de ordin metodologic, trebuie avută grijă şi de aspectele statistice. Tabelul 3.5 conţine o sinteză a celor mai des întâlnite riscuri în cazul validităţii statistice, atât prin impactul pe care îl au asupra puterii statistice (scad puterea statistică), cât şi prin riscul de distorsiune legat de magnitudinea efectului observat (de exemplu, subestimarea efectului).

Tabelul 3.5. *Riscuri privitoare la validitatea statistică*

Riscuri privitoare la validitatea statistică prin impactul lor asupra puterii statistice şi a acurateţei magnitudinii efectului
Restricţia de amplitudine *Exemplificare*: Într-un studiu ce vizează relaţia dintre conştiinciozitate şi performanţa angajaţilor din spital se observă absenţa corelaţiei dintre cele două variabile. La o analiză mai atentă se observă restricţii de amplitudine în cazul ambelor variabile. De exemplu, deşi scala de măsurare a conştiinciozităţii lua valori între 0 (nivel foarte scăzut) şi 20 (nivel extrem de ridicat al conştiinciozităţii), rezultatele colectate de la participanţi au conţinut scoruri cuprinse între 13 şi 18 puncte, astfel că toţi subiecţii testaţi au un nivel mediu superior al conştiinciozităţii). În mod similar, deşi existau 5 trepte de clasificare a performanţei (de la 1 [mult sub aşteptări] la 5 [mult peste aşteptări]) s-a constatat că peste 95% dintre participanţi au primit calificative de 4 sau 5 şi niciunul dintre candidaţi nu a primit un scor de 1 sau 2.

O astfel de situaţie, în care există o variabilitate scăzută a scorurilor, poartă numele de restricţie de amplitudine şi se soldează cu o diminuare puternică a indicatorului statistic calculat (de exemplu, coeficienţi scăzuţi de corelaţie), fapt ce contribuie la apariţia unui risc crescut de a comite o eroare de tip II, respectiv la subestimarea legăturii dintre variabile. Efectul negativ apare atunci când una sau amândouă dintre variabilele analizate suferă de restricţie de amplitudine.
Tot în cadrul restricţiei de amplitudine aş aminti despre efectul de podea (*floor effect*), respectiv de tavan (*ceiling effect*). O ilustrare a efectului de tavan ar fi propunerea de a testa eficienţa unei metode noi de predare a matematicii faţă de metoda clasică la două clase speciale de matematică (cu un nivel mediu al performanţei la matematică la pretest de 9,20 în clasa A, respectiv de 9,25 în clasa B). Chiar dacă metoda ar fi una benefică, nu există un spaţiu mare de creştere, astfel încât o posibilă creştere cu 50 de sutimi (de la 9,20 la 9,70) ar fi insuficientă pentru a obţine rezultate semnificative statistic, în ipoteza menţinerii aceluiaşi scor de 9,25 pentru clasa B). Dacă nivelul de la care s-ar fi pornit ar fi fost unul mai scăzut (de exemplu, o medie de 7,50), ar fi fost posibilă o creştere mai mare după implementarea noii metode (de exemplu, de 1,50 puncte), suficientă pentru a conduce, de această dată, la diferenţe semnificative între grupe.
Efectul de podea se referă la o situaţie opusă. De exemplu, într-un studiu se urmăreşte eficienţa unei tehnici de diminuare a agresivităţii, testată pe participanţi cu agresivitate scăzută (marja de scădere este prea mică pentru a sesiza efectul).

Consistenţa internă scăzută
Exemplificare: Într-un studiu se compară dacă există diferenţe între grupul experimental şi lotul de control cu privire la nivelul anxietăţii sociale. Această variabilă a fost măsurată printr-o scală a cărei consistenţă internă este scăzută (alfa Cronbach fiind 0,46). Restul aspectelor implicate în cercetare au fost implementate corect.
Situaţiile în care măsurătorile esenţiale pentru testarea ipotezei (cum este cazul VD în ipoteze ce vizează comparaţii între grupe) au niveluri scăzute ale consistenţei interne, ne îndeamnă să fim extrem de precauţi şi să avem o încredere scăzută în acurateţea scorurilor obţinute, din cauza erorii de măsurare ridicate. Din punct de vedere statistic, situaţia conduce la creşterea riscului de apariţie a erorii de tip II şi la subestimarea legăturii dintre variabile.

Încălcarea asumpţiilor unui test statistic
Exemplificare: Într-un studiu se urmăreşte dacă există o legătură între venitul lunar al participanţilor (exprimat în lei) şi atitudinea oamenilor faţă de demersul de suspendare a preşedintelui Traian Băsescu evaluată printr-o scală Likert în 7 trepte (1 – total împotrivă, 7 – sprijin total). Datele colectate de la 300 de participanţi (deşi acoperă întreaga plajă de răspunsuri pentru a evita restricţia de amplitudine), indică faptul că cele două variabile au o distribuţie puternic asimetrică (de exemplu, 90% dintre participanţi au declarat venituri până în 2.500 lei net, iar atitudinea privind demersul de suspendare era mai degrabă favorabilă, media fiind plasată la 5,75 în scala de 7 puncte).
În exemplul descris, corelaţia calculată pe aceste date brute va fi atenuată, conducând la efecte similare cu cele menţionate în primele două situaţii de mai sus. Acest caz ilustrativ este numai

unul din zecile de exemple posibile cu privire la situaţia de încălcare a asumpţiilor asociate unui anumit test statistic. De exemplu, în cazul corelaţiei Pearson, una dintre cele mai utilizate tehnici statistice, între asumpţiile de bază pentru a ajunge la un rezultat corect se află: (a) existenţa unei relaţii liniare între variabile; (b) aspectul homoscedastic al asocierii (omogenitatea relaţiei dintre variabile); (c) absenţa devierilor grave de la cazul unei distribuţii normale pentru variabilele analizate; (d) absenţa cazurilor extreme (*outliers*). Pentru mai multe detalii despre acest subiect, vezi Sava (2004/2011).

Aplicarea unui test statistic prin ignorarea asumpţiilor (condiţiilor necesare aplicării acestuia) creşte riscul de a comite atât erori de tip I, cât şi erori de tip II, în funcţie de problema ignorată. Soluţia pentru această problemă constă fie în corectarea problemelor respective acolo unde este posibil (de exemplu, transformarea scorurilor brute pentru a ameliora problemele de distribuţie anormală – vezi partea de final a capitolului 3 în Sava, 2004/2011), fie în aplicarea altor tehnici statistice.

Aglutinarea datelor

Exemplificare: Într-un studiu privind relaţia dintre nivelul de extravertire şi nevoia de senzaţii tari sunt colectate date numerice despre extraversiune printr-o scală de profil ce permite amplitudini ale scorurilor între 0 şi 36. Deşi ar fi putut testa direct ipoteza vizată, cercetătorul decide să împartă participanţii testaţi în introvertiţi, respectiv extravertiţi, în baza mediei (cei cu scoruri sub medie fiind consideraţi introvertiţi, iar cei cu scoruri peste medie, extravertiţi).

Acest gen de transformări, alături de alte variante de aglutinare (de exemplu, realizarea analizelor statistice nu pe baza scorurilor brute, ci pe baza scorului clasei obţinut în baza unui etalon – celebră fiind etalonarea în 5 clase: de la 1 – foarte scăzut, la 5 – foarte ridicat) nu sunt indicate deoarece se pierde o informaţie esenţială, ce poate conduce la distorsionarea rezultatelor şi la efecte specifice restricţiei de amplitudine. De exemplu, dihotomia rezultată în baza unui examen – admişi şi respinşi – pierde din vedere diferenţele intragrup (există o distanţă calitativă mai mare între un individ admis cu 5,01 faţă de un coleg notat cu 10,00, decât faţă de un alt coleg din celălalt grup, al respinşilor, dar notat cu 4,95). Shadish *et al.* (2002) tratează această situaţie ca un caz special al restricţiei de amplitudine.

Raportarea selectivă a rezultatelor

Exemplificare: Într-un studiu ce urmăreşte să identifice dovezi cu privire la validitatea predictivă a unei probe de personalitate construite pe modelul Big-Five sunt analizate corelaţiile dintre cele 5 dimensiuni de personalitate şi un set de 100 de comportamente (câte 20 asociate fiecărei dimensiuni, notate de la 1 [deloc] la 4 [frecvent]). În urma prelucrării statistice a rezultatelor, se constată că doar în 15 cazuri din cele 100 de comportamente, rezultatele indică asocieri semnificative cu dimensiunile respective în direcţia aşteptată. În redactarea articolului, cercetătorul ia decizia de a include doar 20 de comportamente (câte 4 pentru fiecare dimensiune, între care se regăsesc toate cele 15 pentru care s-au obţinut asocieri semnificative statistic), ignorându-le pe celelalte 80 de comportamente, al căror rezultat era nesemnificativ statistic.

În principiu, limitarea la prezentarea numai a acelor rezultate care spun o poveste interesantă este o practică încurajată chiar de editorii revistelor de specialitate, din lipsă de spaţiu editorial. Cu toate acestea, eliminarea selectivă a rezultatelor care nu convin distorsionează clar concluziile ce pot fi trase din studiu. Prezentarea tuturor rezultatelor obţinute ar fi preferabilă pentru a ne forma o impresie realistă, corectă, asupra rezultatelor obţinute, dar şi pentru a aprecia riscul de apariţie a erorii setului de comparaţii (vezi discuţia despre acest tip de eroare la începutul capitolului 4 din Sava, 2004/2011) pentru a aprecia care este probabilitatea de a obţine din întâmplare rezultate semnificative statistic, în contextul în care doar 15% dintre comportamente evaluate au fost asociate unui rezultat semnificativ statistic. Dacă prezentarea tuturor rezultatelor nu este posibilă din motive editoriale, preferabil ar fi să se apeleze la strategii complementare. De exemplu, dacă fiecare dimensiune de personalitate era asociată cu un set de 20 de comportamente, se poate pune accent pe prezentarea rezultatelor dintre relaţia dintre dimensiunea respectivă şi un scor compozit (global) al comportamentelor, cum ar fi media comportamentelor sau scorul factorial obţinut în urma includerii acestora într-o analiză factorială (a componentelor principale). Un avantaj suplimentar al acestei decizii este evitarea cazurilor de restricţie de amplitudine (de exemplu, în cazul însumării celor 20 de comportamente rezultatul variabilei în cauză variază teoretic între 20 şi 80, în timp ce în analizele bazate pe fiecare comportament în parte, variază teoretic doar între 1 şi 4).
Eterogenitate la nivel intragrupal *Exemplificare*: Într-un studiu interesat de stabilirea eficienţei unui program de slăbire cu ajutorul unor exerciţii de imagerie mentală, nu s-au sesizat diferenţe semnificative între grupul experimental (m = 72 kg, as = 30 kg) şi grupul de control (m = 80 kg, as = 28 kg). În schimb, un alt studiu pe aceeaşi temă identifică diferenţe semnificative statistic, sprijinind eficienţa intervenţiei (m = 72 kg, as = 14 kg în grupul experimental, respectiv m = 80 kg şi as = 12 kg în grupul de control). Deşi diferenţa dintre mediile celor două grupe este aceeaşi în cele două studii, rezultatele sunt diferite! Eşecul primului studiu de a găsi diferenţe semnificative statistic este determinat de o eterogenitatea prea mare a participanţilor din grupele comparate, fapt reflectat prin niveluri ridicate ale abaterilor standard. Astfel, este mai simplu de detectat o diferenţă de 10 kg între mediile celor două grupe dacă acestea sunt alcătuite din indivizi relativ omogeni (de exemplu, cântărind între 85 şi 95 de kg) decât din indivizi cu o greutate eterogenă (de exemplu, cântărind între 70 şi 120 de kg). Morala: aşa cum se vorbeşte popular despre un colesterol bun şi unul rău, la fel am putea afirma în cazul împrăştierii datelor. Aceasta poate fi benefică (în studii corelaţionale sau dacă vorbim despre dispersia intergrup – diferenţele dintre grupe) sau nocivă pentru validitatea statistică a studiului (cazul dispersiei intragrup – diferenţe mari între indivizii cuprinşi în acelaşi grup). Eterogenitatea poate viza, fie variabila dependentă, fie variabile cu potenţial confundat, fie chiar modalităţile variabilei independente, cum este cazul implementării unei intervenţii nestandardizate.
Apelul la tehnici statistice inadecvate

Exemplificare: Într-un studiu s-a urmărit dacă există o asociere între nivelul de inteligenţă (operaţionalizat prin scorul IQ) şi venitul lunar (operaţionalizat în lei net/lună). Pentru aceasta, cercetătorul a apelat la tehnica hi pătrat al gradului de omogenitate şi a concluzionat că nu există o asociere semnificativă statistic.
Am lăsat la final, poate cel mai evident risc privitor la validitatea statistică, aceea de a aplica o tehnică statistică greşită (nepotrivită), în cazul descris tehnica corectă fiind coeficientul de corelaţie lineară Bravais-Pearson. Deşi asemenea cazuri sunt extrem de rare în literatura de specialitate publicată în reviste cu prestigiu internaţional, am amintit de aceast risc pentru că, din păcate, sunt încă întâlnite ocazional în România, inclusiv în rândul unor lucrări de doctorat sau al unor articole publicate în reviste de specialitate din ţară.

3.4.2. Validitatea internă

Demonstrarea din punct de vedere statistic a unei asocieri între două variabile de interes este o condiţie insuficientă pentru a putea infera relaţii cauzale. Cu alte cuvinte, nu orice corelaţie semnificativă statistic este un rezultat al prezenţei unei relaţii de tip cauză-efect. Pentru a demonstra existenţa unei asemenea relaţii cauzale, (a) pe lângă demonstrarea unei asocieri semnificative statistic, (b) ar trebui să se constate că modificările în variabila cauză precedă temporal modificările observate în variabila efect, (c) în condiţiile în care sunt eliminate alte explicaţii alternative care ar putea fi responsabile pentru legătura respectivă (Shadish *et al.*, 2002).

Dacă problema obţinerii unui indicator precis pentru relaţia statistică dintre cele două variabile este una de validitate statistică, iar aspectul temporalităţii (ordinea în apariţia modificărilor) poate fi obţinut destul de simplu prin design[7] sau prin analiză logică[8], problema eliminării explicaţiilor alternative pentru a păstra doar explicaţia cauzală drept sursă a legăturii dintre variabile este o problemă de validitate internă. De multe ori este posibil ca demersul statistic să fie excelent realizat, însă să existe erori în raţionamentele cauzale utilizate pentru acest lucru.

Pentru a ilustra această stare de fapt voi apela la un studiu cvasiexperimental ipotetic cu design pretest-posttest, fără grup de control. Într-o iarnă, un cercetător este interesat să propună un program bazat pe exerciţii Yoga pentru a creşte imunitatea organismului (operaţionalizat prin diminuarea numărului de răceli). Pentru aceasta apelează la un lot de 100 de voluntari pe care îi învaţă şi îi monitorizează timp de şase luni în practicarea exerciţiilor Yoga. Din punct de vedere statistic, cercetătorul constată un număr semnificativ mai redus de răceli în ultima lună de monitorizare a exerciţiilor comparativ cu situaţia iniţială, ce viza frecvenţa răcelilor în luna premergătoare începerii programului de antrenament. Este acest rezultat suficient pentru a valida inferenţa cercetătorului, că prin antrenament Yoga se reduc problemele de sănătate, că exerciţiile sunt benefice pentru imunitatea organismului? Răspunsul, din perspectiva validităţii interne a acestui studiu, este desigur unul negativ *atâta timp cât pot fi oferite explicaţii alternative plauzibile* pentru rezultatul obţinut. De pildă, o explicaţia alternativă pentru diminuarea numărului de răceli poate fi dată de aspectul sezonier al epidemiilor de răceală şi gripă (mai des întâlnite iarna – momentul evaluării pretest, decât vara – momentul evaluării posttest). O altă explicaţie posibilă vizează alimentaţia. Este posibil ca odată cu practicarea exerciţiilor Yoga, participanţii să facă schimbări în tipul de alimentaţie folosit şi să înceapă să mănânce mai sănătos. Prin urmare, ameliorarea observată la nivelul frecvenţei îmbolnăvirilor poate fi cauzată de noul regimul alimentar sănătos şi nu de exerciţiile Yoga. De asemenea, rezultatul poate fi o consecinţă a distorsiunii în selecţia participanţilor. Este foarte posibil ca participanţii care au acceptat să fie voluntari în acest studiu cu o durată de şase luni să înceapă să manifeste o preocupare deosebită pentru sănătatea lor, ca urmare a unui istoric personal asociat unor îmbolnăviri frecvente.

De aceea este posibil să acorde mai multă grijă menţinerii sănătăţii (de exemplu, prin consum zilnic de vitamina C, alimentaţie variată, program regulat de somn etc.), iar această modificare să fie cauza diminuării frecvenţei răcelilor înregistrate.

3.4.2.1. Problema celei de-a treia variabile

Aceste exemple subliniază un element comun şi esenţial pentru problematica validităţii interne a unui studiu, denumită generic de Brewer (2000) drept „problema celei de-a treia variabile" şi fiind sursa reală a relaţiei observate dintre VI şi VD. Deşi, în principiu, validitatea internă ar trebui să fie o problemă specifică doar pentru studiile care urmăresc un obiectiv de tip cauzal (cele longitudinale, cele cvasiexperimentale şi cele experimentale), ea poate fi adresată şi în alte tipuri de studii, dacă cercetătorii fac inferenţe de natură cauzală cu privire la rezultatele obţinute. Bunăoară, prezentarea datelor provizorii pentru recensământul din 2011 şi discursul public însoţitor pentru a explica rezultatele obţinute constituie o sursă bogată de discuţie din perspectiva celei de-a treia variabile. Spre exemplu, s-a observat că există o asociere între judeţele care au cunoscut cea mai mare scădere demografică faţă de recensământul din 2002 şi existenţa unor condiţii igienice adecvate (cum ar existenţa unei băi în locuinţă). Astfel, în judeţele a căror populaţie a scăzut cel mai mult se observă un procent mai scăzut de gospodării dotate cu baie în casă. Din punct de vedere logic, este puţin probabil ca această asociere să indice o relaţie cauzală între cele două. Mai mult, din punct de vedere metodologic este imposibil să fie valide asemenea inferenţe cauzale, fiind vorba de un studiu descriptiv transversal, inadecvat pentru a răspunde la

obiective de tip cauzal (deoarece nu poate adresa condiţiile b şi c necesare pentru a considera o relaţie dintre variabile ca fiind una cauzală).

Desigur, prin implicarea în ecuaţie a altor variabile este posibil să se ofere unele explicaţii (provizorii). De exemplu, este posibil ca ambele variabile să reprezinte doar efecte ale unei cauze comune. Aceasta ar putea fi un nivel mai mare de sărăcie în judeţele respective sau un procent mai ridicat al populaţiei cu rezidenţă în mediul rural etc. Pentru că atât sărăcia, cât şi rezidenţa în mediul rural pot fi facilitatori ai existenţei unor băi în afara locuinţei şi ai mobilităţii demografice. Acest tip de relaţie, cum ar fi cea dintre pierderea populaţiei din judeţe şi absenţa băilor în locuinţă poartă denumirea în statistică de legătură falsă (*spurious correlation*), fiind prezentată în partea dreaptă a figurii 3.3.

Complexitatea problemei celei de-a treia variabile este descrisă în figura 3.3. De exemplu, dacă ne raportăm la partea stângă a figurii şi dacă notăm cu X antrenamentul Yoga, cu Y frecvenţa răcelilor şi cu C alimentaţia sănătoasă, putem observa că inferenţa „Yoga ajută la diminuarea frecvenţei răcelilor” nu reprezintă o cauza reală, deşi statistic poate exista o legătură între ele. Se ajunge la această afirmaţie eronată, pentru simplul fapt că omitem din analiză asocierea dintre Yoga şi alimentaţia sănătoasă, aceasta din urmă fiind cauza probabilă pentru îmbunătăţirea stării de sănătate. În această situaţie vorbim tot despre existenţa unei legături false între X şi Y, determinate de prezenţa unei a treia variabile C, care joacă rolul unei variabile confundate.

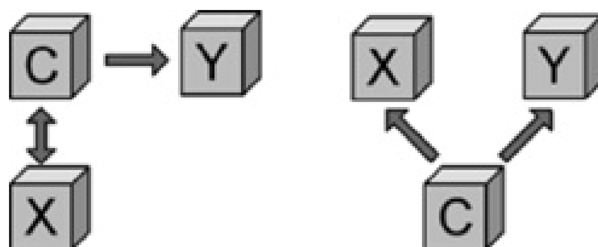


Figura 3.3. *Situaţia unor corelaţii false*

Teoretic, soluţia metodologică pentru situaţia corelaţiilor false ar fi aceea de a varia nivelul variabilei independente şi a urmări efectul acesteia asupra variabilei dependente, în contextul în care s-ar controla (ţine constantă influenţa) nivelul variabilei terţe (confundate).

Pe lângă problematica corelaţiilor false, în cadrul discuţiei despre cea de-a treia variabilă pot fi aduse în discuţie relaţiile de mediere, respectiv de moderare. Păstrând în minte exemplul utilizării exerciţiilor Yoga pentru a reduce frecvenţa răcelilor, este posibil de luat în calcul un alt scenariu interpretativ. De exemplu, este posibil ca între practicarea Yoga şi regimul alimentar sănătos să nu existe doar o relaţie simplă de asociere (o parte însemnată dintre practicanţii Yoga mănâncă sănătos şi se confundă efectele determinate de acestea asupra unei variabile terţe), ci să existe o relaţie cauzală în sine (studii experimentale bine controlate, în care se constată că practicanţii Yoga, spre deosebire de voluntarii aflaţi pe lisa de aşteptare pentru a începe o formare în Yoga, mănâncă mai sănătos). În acest caz, lanţul cauzal devine următorul: practicarea Yoga contribuie la îmbunătăţirea regimului alimentar al practicanţilor, iar un regim alimentar corespunzător diminuează problemele de sănătate. Într-un asemenea scenariu, deşi se poate spune că exerciţiile Yoga constituie o cauză indirectă a reducerii frecvenţei răcelilor, explicaţia pentru acest lucru nu vizează exerciţiile fizice în sine, ci un efect al filosofiei Yoga, regimul alimentar sănătos. Gradul de validitate al unui asemenea rezultat ipotetic este apreciat diferit. Din punctul de vedere al unui studiu care urmăreşte un obiectiv cauzal-descriptiv, legătura dintre practicarea Yoga şi ameliorarea stării de sănătate ar fi o inferenţă validă, pentru că în acest caz nu contează şi validarea explicaţiei pentru relaţia cauzală constatată, ci numai validarea existenţei relaţiei cauzale. În schimb, din perspectiva

unui studiu care urmăreşte un obiectiv cauzal-explicativ, inferenţa ar fi validă doar în contextul menţionării întregului lanţ cauzal şi a existenţei unui sprijin empiric pentru aceasta prin datele colectate. În această situaţie, suntem interesaţi atât de validitatea inferenţelor cu privire la relaţia cauzală în sine, cât şi de validitatea inferenţelor cu privire la explicaţia găsită pentru relaţia cauzală respectivă.

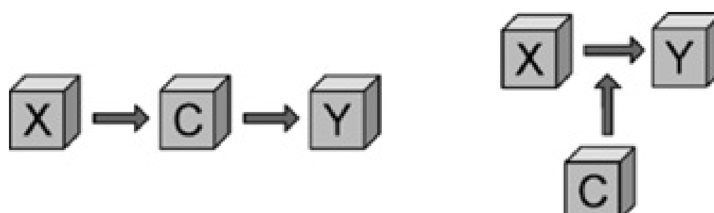


Figura 3.4. *Cazul unei relaţii de mediere (stânga), respectiv al relaţiei de moderare (dreapta)*

În sfârşit, o relaţie cauzală poate fi înţeleasă greşit şi în situaţia în care cea de-a treia variabilă joacă un rol moderator extrem. Prin variabilă moderatoare se înţelege acea variabilă care afectează intensitatea relaţiei dintre două variabile. În cazuri extreme se poate ajunge la situaţia ciudată de a se inversa chiar efectul observat. De exemplu, impactul pe care muzica clasică îl are asupra relaxării psihice este moderat de nivelul de educaţie sau de modelul cultural adoptat. Pentru persoane cu studii superioare, cu un nivel ridicat de cultură generală este foarte probabil ca muzica clasică să faciliteze relaxarea. În schimb, pentru persoane cu un nivel de şcolarizare scăzut, proveniţi dintr-un mediu cultural în care se preferă ascultarea de manele, există probabilitatea ca muzica clasică nu numai să nu funcţioneze în a induce relaxare, ci chiar să conducă la un efect invers, de creştere a nivelului de excitare (iritare, încordare). Dacă studiul s-ar derula pe participanţi eterogeni, aparţinând ambelor culturi, s-ar putea ca relaţia cauzală să treacă neobservată, cele două direcţii anulându-se reciproc. În consecinţă, este posibil ca ignorarea unei

variabile terţe cu rol moderator să conducă la concluzii distorsionate cu privire la relaţia cauzală dintre variabilele asupra cărora se produce efectul de moderare.

Problematica variabilelor confundate este esenţială din perspectiva gradului de validitate al inferenţelor de tip cauzal. Tabelul 3.6 conţine cele mai relevante riscuri care au drept cauză impactul variabilelor confundate asupra VD (Shadish *et al.*, 2002; Cook şi Campbell, 1979).

Tabelul 3.6. *Riscuri privitoare la validitatea internă*

Istoricul *Exemplificare*: Într-un studiu experimental sunt comparate două grupe de elevi de gimnaziu, distribuiţi aleatoriu, fie în grupa experimentală care urmează un program de consiliere de tip *peer to peer* (consiliere colegială), fie în grupa de control (lista de aşteptare). Scopul studiului este de a stabili eficienţa consilierii colegiale în prevenirea fumatului. Pe durata derulării experimentului, şcoala din care fac parte elevii din studiu este selectată într-un program naţional educativ despre consecinţele negative ale consumului de alcool şi tutun. La finalul derulării intervenţiei nu sunt identificate diferenţe semnificative statistic între cele două grupe. Acest tip de risc se referă la orice tip de eveniment neplanificat care afectează evoluţia variabilei dependente, astfel încât nu se poate aprecia cu certitudine dacă efectul observat este datorat intervenţiei dorite (consilierii colegiale) sau evenimentului (istoricului) neplanificat (programul educaţional cu rol informativ). Pentru a evita asemenea situaţii, se recomandă includerea în formularul de consimţire a unei solicitări exprese ca, în cazul în care acceptă să participe în studiu, să nu urmeze alte tratamente alternative. Aspectele de istoric însă pot privi orice situaţie apărută pe durata implementării intervenţiei, inclusiv: (a) situaţii cu impact emoţional major asupra individului cuprins în studiu (cum ar fi situaţiile de deces în familie sau aflarea faptului că suferă de o boală gravă, când intervenţia este menită să reducă depresia); (b) evenimente promovate intens în mass-media (rata scăzută de promovabilitate la bacalaureat în mijlocul unui studiu adresat elevilor de liceu, care îşi propune să testeze eficienţa unei noi metode de învăţare) etc. În sfârşit, deşi când ne referim la istoric, avem în vedere un impact generalizat asupra tuturor participanţilor la studiu, pot fi analizate şi situaţii individuale, pentru a decide dacă individul poate fi păstrat ca subiect în studiu (de exemplu, se îmbolnăveşte şi nu poate participa decât la 30% dintre şedinţele planificate).

Soluţii uzuale: Utilizarea unui grup de control (evenimentele cu caracter public vor afecta la fel de mult grupul experimental şi pe cel de control) şi debriefing sau interviuri individuale efectuate la final, după caz (pentru a detecta situaţii încadrabile ca istoric, la nivel individual).

Maturarea

Exemplificare: Într-un studiu cvasiexperimental de tip pretest-posttest fără grup de control au fost urmăriţi 20 de indivizi ce sufereau de oboseală cronică. Aceştia au fost testaţi cu privire la nivelul lor de oboseală atât înainte, cât şi după derularea unui program de antrenare a gândirii pozitive, derulat pe o perioadă de şase săptămâni. La finalul intervenţiei se observă o scădere semnificativă statistic a nivelului de oboseală.

Acest tip de risc se referă la faptul că unele dintre evenimentele studiate se pot ameliora de la sine. Astfel, multe tulburări se pot remite spontan. De asemenea, în anumite situaţii, deficienţele

se corectează de la sine odată cu maturizarea sistemului nervos central (vezi diminuarea simptomatologiei asociate enurezisului sau deficitului de atenţie etc., odată cu înaintarea în vârstă). La fel de adevărat este faptul că, în cazul situaţiilor cu o evoluţie ciclică (sezonieră), se poate ajunge şi la reversul situaţiei, acelea de a subestima efectul intervenţiei. Prin urmare, problema cu acest tip de risc este faptul că nu vom putea şti cu certitudine dacă modificările sunt cauzate de intervenţia urmărită sau sunt o consecinţă a maturării. Pentru a elimina această risc este imperios să se utilizeze un grup de control care să nu beneficieze de intervenţie. Acest tip de risc apare frecvent în studiile de psihologia dezvoltării, îndeosebi la copiii de vârste mici, unde inclusiv cei din grupul de control au multe achiziţii psihice într-un interval scurt de timp.

Soluţii uzuale: Utilizarea unui grup de control echivalent ca vârstă sau alte caracteristici relevante pentru maturare (prin această soluţie, maturarea va afecta la fel de mult grupul experimental şi pe cel de control, fiind controlat astfel impactul maturării asupra VD).

Testarea

Exemplificare: În cazul în care participanţii la studiu sunt supuşi unor testări repetate (de exemplu, sunt rugaţi să completeze *aceeaşi* scală de inteligenţă generală de trei ori într-un interval de două luni, în cadrul unui program menit să augmenteze IQ-ul unui individ): există riscul unei distorsiuni prin faptul că, devenind familiari cu proba respectivă au şanse să obţină un punctaj mai bun pe măsură ce repetă testul (prin învăţarea sau reamintirea unor răspunsuri).

În acest tip de risc nu se modifică instrumentul, ci raportarea şi modul de răspuns al individului la instrumentul de evaluare utilizat repetat. Riscul este activ pentru probe subiective (evaluări prin autoraportare sau heteroraportare), fiind cu atât mai probabilă cu cât există un număr mai mare de repetări, într-un interval mai scurt de timp între şedinţele de evaluare.

Soluţii uzuale: Utilizarea ca VD a unor măsurători obiective sau extinderea distanţei dintre şedinţele de evaluare. Acolo unde există posibilitatea, pot fi utilizate probe paralele.
Instrumentarea *Exemplificare*: Participanţii la studiu longitudinal, întins pe o perioadă lungă de timp (de exemplu, 15 ani, cu evaluări de la 4 la 19 ani) sunt examinaţi prin trei scale de inteligenţă diferite (una specifică copiilor preşcolari [4-6 ani], una specifică elevilor de gimnaziu [7-14 ani] şi una destinată evaluării adolescenţilor). Scopul studiului este de a evidenţia dacă un program menit să augmenteze IQ-ul unui individ este eficient. Rezultatele obţinute la finalul celor 15 ani de urmărire nu identifică diferenţe semnificative statistic între grupul experimental şi cel de control. Prin acest tip de risc (specific mai degrabă studiilor longitudinale derulate pe o perioadă mai lungă de timp) ne referim, în principal, la modificări în instrumentele utilizate (de exemplu, proba de măsurare a IQ-ului la 4 ani nu e potrivită pentru a fi utilizată pe subiecţi de 19 ani, deoarece ar fi prea uşoară, prin urmare nediscriminativă). Tot în această categorie intră şi situaţia utilizării unor versiuni diferite ale aceluiaşi test. De exemplu, este posibil ca un test să fie revizuit şi actualizat în timp din raţiuni psihometrice sau de modificare a conţinutului conceptului pe care îl vizează. Astfel, versiunea aplicată iniţial (la începutul studiului) poate fi diferită de cea revizuită, aplicată la finalul studiului. Pentru a evita asemenea situaţii, ar trebui ca instrumentele diferite utilizate să fie testate din perspectiva invarianţei lor metrice. Un alt caz de instrumentare apare în cazul utilizării unor metode apreciative (subiective) de cotare, cum ar fi cazul observaţiei. De exemplu, poate interveni oboseala care afectează corectitudinea înregistrărilor sau pot fi implicaţi asistenţi de cercetare diferiţi de la un moment la altul al testării, fiecare cu un stil personal de apreciere a evoluţiei participanţilor. În mod similar, poate interveni o diminuare a exigenţei în aprecierea subiectivă a comportamentelor, în direcţia aşteptată prin ipoteze, îndeosebi în contextul studiilor transparente, unde nu este posibilă mascarea tipului de intervenţie primită.
Soluţii uzuale: Stabilirea nivelului de invarianţă dintre instrumente (în cazul probelor standardizate revizuite); apelul la probe standardizate şi/sau obiective; mascarea intervenţiei prin designuri de tip simplu-orb sau dublu-orb (asistenţii de cercetare să cunoască din ce condiţie experimentală face parte individual evaluat) etc.
Regresia către medie *Exemplificare*: Într-un studiu efectuat în cadrul unui liceu cu multiple probleme de disciplină, psihologul şcolar aplică un test de măsurare a nivelului de agresivitate, în baza căruia selectează 20 de băieţi – cei care înregistrau cel mai ridicat nivel de agresivitate. Cu acordul acestora aplică un program de consiliere bazat pe tehnici comportamentale în scopul reducerii agresivităţii acestora. Rezultatul statistic obţinut indică eficienţa programului, nivelul de

agresivitate măsurat postintervenţie fiind semnificativ statistic mai scăzut decât la momentul pretestului.

Acest tip de risc este unul subtil pentru cei mai puţin familiarizaţi cu principiile statisticii. Regresia către medie se referă la probabilitatea scăzută ca la o a doua testare să se obţină tot scoruri extreme. Astfel, cei cu scoruri foarte mari, vor avea şanse mari să obţină scoruri ceva mai scăzute, iar cei cu scoruri foarte mici, vor avea şanse mari să obţină scoruri ceva mai ridicate decât la pretest. Din punctul de vedere al practicianului, acest efect poate fi o mană cerească. Astfel, este foarte probabil ca un pacient să apeleze la psihoterapie când nivelul său de distres este unul maxim sau ca o companie să apeleze la serviciile unui consultant atunci când problema „arde". Chiar dacă propunerile practicianului nu sunt unele eficiente, este posibil să se observe o diminuare (e drept nu foarte mare) a intensităţii problemei respective, ceea ce poate induce impresia greşită că soluţiile oferite au ajutat la reducerea gravităţii problemei.

Soluţii uzuale: În cazul studiilor experimentale ar fi de dorit să se selecteze un număr suficient de mare de persoane cu scoruri extreme (dacă populaţia evaluată este suficient de mare), astfel încât participanţii selectaţi să fie distribuiţi randomizat fie în grupul experimental, fie în cel de control. Această măsură ar conduce la egalizarea efectului regresiei către medie în cele două grupe, controlând astfel impactul acestui posibil risc. În cazul studiilor cvasiexperimentale fără grup de control, soluţia ar fi ca participanţii selectaţi în studiu să facă parte în număr egal din ambele extreme (cu agresivitate foarte ridicată, respectiv cu agresivitate foarte scăzută).

Distorsiunea în selectarea participanţilor

Exemplificare: Într-un studiu cvasiexperimental, un consilier familial doreşte să testeze eficienţa tehnicii de respiraţie Lamaze în reducerea durerilor din timpul travaliului naşterii naturale. În urma sprijinului logistic primit din partea unei clinici de obstetrică şi ginecologie, consilierul organizează o întâlnire informativă pentru viitoarele mame la care participă 50 de persoane. În urma întâlnirii respective recrutează 23 de persoane care s-au arătat interesate şi încântate de metodă şi au fost de acord să participe la un număr de 10 şedinţe destinate deprinderii tehnicii Lamaze. Celelalte 27 de persoane care nu au dorit, din diferite motive, să înveţe tehnica respectivă au fost de acord să participe la studiu ca grup de control.

Problema cu acest tip de risc[9] este că cele două grupuri pot fi diferite încă de la început. În acest caz, participantele din grupul experimental au un nivel diferit de motivare şi interes faţă de metoda evaluată comparativ cu grupul de control, iar rezultatul favorabil tehnicii ar putea fi cauzat de acest plus motivaţional şi nu de caracteristicile tehnicii în sine. Sau este posibil ca în grupul experimental să fi fost incluse mai multe viitoare mame cu sensibilitate mare la durere (speriate de gândul naşterii naturale), fapt care ar impieta găsirea unor efecte în favoarea eficienţei tehnicii Lamaze, pentru simplul fapt că există diferenţe între cele două grupe în ceea ce priveşte sensibilitatea faţă de durere.

Riscul apare în orice context în care distribuirea participanţilor în condiţia experimentală alocată nu se face în mod aleatoriu. De exemplu, testarea eficienţei unei metode noi de predare comparativ cu cea clasică, în două clase paralele, a V-a A – primind metoda nouă, iar a V-a B – primind metoda clasică, nu poate fi făcută adecvat. Eventualele diferenţe sesizate la final pot fi explicate atât prin plusul adus de metoda nouă, cât şi prin diferenţele naturale existente între cele două clase încă înainte de a derula studiului (cum ar fi cazul în care sunt mai mulţi elevi buni în clasa a V-a A decât în clasa a V-a B).
Soluţii uzuale: Apelul la studii experimentale, care prin definiţie presupun alocarea randomizată a participanţilor în grupele experimentale (pentru detalii despre implementarea corectă a randomizării, vezi secţiunea 4.1.3).

Retragerea diferenţiată a participanţilor
Exemplificare: Un director de programe educaţionale este interesat să observe care dintre cele două discipline facultative disponibile conduce la un grad mai mare de mulţumire faţă de opţiunea avută. Astfel, 100 de participanţi la studiu au acceptat să fie distribuiţi randomizat, fie într-un curs de astronomie (N = 50), fie într-un curs de astrologie (N = 50). La finalul cursului, studenţii au fost chestionaţi cu privire la atitudinea lor faţă de cursul urmat (cât de mult le-a plăcut). Au răspuns la chestionarul aplicat la posttest un număr de 25 de studenţi care au urmat până la final cursul de astronomie şi un număr de 45 de studenţi care au urmat cursul de astrologie. Restul studenţilor rămaşi nechestionaţi nu au fost evaluaţi pentru că au renunţat să finalizeze cursul respectiv, din diferite motive. Nu au fost identificate diferenţe semnificative statistic între cele două cursuri, ambele părând a fi la fel de atractive.
Problema ilustrată în exemplul de mai sus este aceea că la final comparaţia s-a făcut numai pe 50% dintre cei care au urmat astronomie şi pe 90% dintre cei care au fost alocaţi cursului de astrologie. Este foarte posibil ca absenţa diferenţei în atractivitatea celor două cursuri să fie cauzată de faptul că la final au fost testaţi doar cei care considerat utile informaţiile predate. Este foarte posibil ca cei 50% dintre cursanţi care au renunţat să finalizeze cursul de astronomie să fi făcut acest lucru pentru că nu au fost atraşi de disciplina respectivă. Dacă ar fi rămas în studiu până la evaluarea finală, este probabil că balanţa s-ar fi înclinat către o mai mare atractivitate pentru cursul de astrologie. De aceea, compararea performanţei înregistrate la nivel de postintervenţie între două grupe cu un nivel diferit de pierdere a participanţilor, semnifică a compara două grupe care nu mai sunt echivalente, în pofida randomizării iniţiale.
Pe lângă pierderea diferenţiată a participanţilor (în limba engleză sunt utilizaţi doi termeni: *attrition* sau *participant mortality*), o problemă, dar nu la fel de gravă, o reprezintă pierderea participanţilor în sine (nu se ştie dacă intervenţia ar mai fi fost eficientă dacă am fi păstrat toţi indivizii incluşi iniţial în studiu, inclusiv pe aceia care nu s-au regăsit/nu au perceput utilitatea intervenţiei primite). În domeniul psihoterapiei, un procent de 15-20% de participanţi care renunţă să finalizeze întreg programul convenit

poate fi considerat acceptabil, fiind necesar să se ţină seama de acest lucru atunci când se stabileşte mărimea eşantionului utilizat în studiu (fiind incluşi cu 15-20% mai mulţi participanţi decât ar fi necesar).
Soluţii uzuale: Utilizarea, acolo unde este posibil acest lucru, a unui grup de control care să primească o intervenţie dovedită a fi eficientă în locul unui lot de control de tipul absenţei intervenţiei sau a listei de aşteptare.

Tabelul 3.7. *Riscuri privitoare la validitatea internă în funcţie de designul utilizat*

	Rezultat obţinut		S Long	Studii cvasiexperimentale			Studii experimentale		
Tip de risc	$p < 0,05$	$p > 0,05$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Istoricul*	X	X	+	+	–	–	–	–	–
Maturarea*	X		+	+	–	–	–	–	–
Testarea*	X	X	+	+	–	–	–	–	–
Instrumentarea	X	X	+	+	–	?	–	–	–
Regresia către medie	X		–	?	–	–	–	–	–
Selecţia distorsionată	X	X	–	–	+	–	–	–	–
Pierderea (diferită)	X	X	+	+	+	+	+	?	–

Legendă: $p < 0,05$ studiul conclude că intervenţia este una eficientă;
$p > 0,05$ studiul conclude că intervenţia nu este eficientă.
* Problema în cazul acestor tipuri de riscuri este dependentă de durata existentă între pretest şi posttest. În cazul istoricului şi al maturării, riscul creşte pe măsură ce se scurge mai mult timp între testări. În cazul testării, direcţia este una opusă, riscul creşte pe măsură ce timpul dintre testări este mai scurt.
(1) – studiu non-experimental longitudinal;
(2) – studiu cvasiexperimental cu pretest/posttest, fără grup de control;
(3) – studiu cvasiexperimental cu pretest/posttest şi grup de control pasiv (nu primeşte nimic);
(4) – studiu cvasiexperimental, designul liniei întrerupte de regresie;
(5) – studiu randomizat (experimental) pretest/posttest, cu grup de control pasiv;
(6) – studiu randomizat pretest/posttest, cu grup de control activ (intervenţie uzuală, intervenţie alternativă);
(7) – studiu randomizat pretest/posttest, cu grup de control activ, cu mascarea intervenţiei primite;
X – riscul respectiv, o posibilă explicaţie pentru rezultatul obţinut;
+ risc important;
? risc posibil, în funcţie de context;

– risc absent sau improbabil.

Tabelul 3.7 poate fi citit în două modalităţi diferite, prin raportare la coloane (rezultatul obţinut, respectiv tipul de studii) sau prin raportare la linii (tipul de risc). De exemplu, dacă luăm în considerare rezultatul studiului, iar acesta indică ineficienţa intervenţiei (nu există diferenţe semnificative statistic), putem fi siguri că maturarea sau regresia către medie nu constituie riscuri pentru acest rezultat. Prin definiţie, acest tip de riscuri conduce la erori de tip I (concluzii că intervenţia este eficientă, deşi nu este adevărat acest lucru, diferenţele semnificative fiind explicate ca urmare a maturării sau a regresiei către medie, acolo unde este cazul). În schimb, dacă rezultatul obţinut indică faptul că intervenţia este una eficientă, observăm că toate tipurile de risc discutate pot interveni activ.

În mod similar, dacă luăm în considerare tipul de design avut în vedere, de exemplu, studiul 3 din tabelul 3.7, reprezentând un design cvasiexperimental cu pretest şi posttest şi cu un grup de control care nu primeşte niciun fel de intervenţie, vom putea desprinde următoarele informaţii legate de riscurile existente. Studiul prezintă riscuri majore îndeosebi cu referire la: (a) distorsiunea în selecţie (în absenţa randomizării fiind posibil ca cele două grupe comparate să difere de la nivelul pretestului, cu privire la nivelul VD sau a altor factori cu impact asupra evoluţiei VD); (b) pierderea diferenţiată a participanţilor, ca urmare a: 1. numărului mai mare de participanţii alocaţi grupului de control retraşi comparativ cu cei din grupul experimental, primii fiind nemulţumiţi că nu primesc o intervenţie pentru problema lor; 2. numărului mai mare de participanţi alocaţi în grupul experimental retraşi comparativ cu cei din grupul de control, ca urmare a unei intervenţii ce necesită un efort prelungit sau care are o multitudine de efecte secundare negative. Aspecte precum istoricul, maturarea, testarea, instrumentarea şi

regresia către medie sunt considerate riscuri improbabile pentru acest tip de design, nu pentru că nu ar fi posibilă apariţia lor, ci pentru că de cele mai multe ori, atunci când ele apar, tind să afecteze la fel de mult ambele grupe din studiu, atât pe cel experimental, cât şi pe cel de control. Prin egalizarea efectelor lor (menţinerea constantă a influenţei lor) am putea urmări în continuare dacă diferenţa existentă între grupul experimental (cel care primeşte intervenţia) şi grupul de control (cel care nu primeşte intervenţia) este una semnificativă statistic. În sfârşit, în cazul unui design de tip cvasiexperimental precum studiul 3, care prin definiţie suferă de distorsiunea în selecţia participanţilor, oricând este posibil să apară interacţiuni între acest risc şi celelalte (vezi nota 1 de la pagina 100).

Tot din analiza pe linii (tipuri de risc) pot fi desprinse o serie de concluzii generale cum ar fi: (a) introducerea unui grup de control reduce foarte mult riscurile de tipul istoricului, maturării, testării, instrumentării sau regresiei către medie prin menţinerea constantă a influenţei acestor factori de risc atât în grupul experimental, cât şi în cel de control; (b) apelul la distribuirea aleatorie a indivizilor în condiţiile experimentale elimină distorsiunea de selecţie, ca posibil risc (cel puţin teoretic, în condiţiile asigurării unui număr suficient de mare de subiecţi incluşi în fiecare condiţie experimentală). Astfel, dacă luăm studiul 6 (probabil cel mai bun design posibil în domeniul psihologiei[10] – un studiu randomizat controlat cu un grup experimental ce beneficiază de o intervenţie inovativă, respectiv un grup de control activ, căruia i se oferă intervenţia standard), vom constata că majoritatea riscurilor privitoare la validitatea internă sunt absente sau improbabile. Rămân mai crescute riscurile pentru pierderea diferenţiată a participanţilor (în cazul în care cele două forme de intervenţie nu sunt la fel de atractive pentru subiecţi sau diferă

semnificativ în ceea ce priveşte cantitatea de efort şi de resurse, inclusiv de timp, implicate).

În încheierea discuţiei despre riscurile privitoare la validitatea internă a unui studiu merită subliniat faptul că toate aspectele menţionate anterior vizează punctul (c) din perspectiva condiţiilor necesare pentru a demonstra o relaţie cauzală: măsurile menite să reducă probabilitatea ca efectul observat să fie cauzat de variabile confundate şi nu de intervenţia urmărită. La aceste riscuri se adaugă cel cu privire la ambiguitatea stabilirii cu claritate a relaţiei temporale. Astfel, în studiile corelaţionale transversale (în care toate măsurătorile sunt colectate în acelaşi timp) este imposibil de decelat dacă modificările în variabila-cauză precedă modificările la nivelul efectelor observate sau dacă situaţia este una inversă. De aceea, numai studiile în care modificările la nivelul presupusei variabile precedă măsurarea efectelor pot fi utile în a adresa întrebări de tip cauzal. În această categorie intră studiile experimentale/cvasiexperimentale în care manipularea VI are loc înainte de a măsura impactul la nivelul VD la momentul postintervenţiei. Tot aici putem include studiile non-experimentale longitudinale (care presupun mai multe valuri de colectare a datelor – cel puţin trei), situaţie în care putem observa dacă modificări în presupusa variabilă cauză de la momentul T0 la T1 precedă temporal şi covariază cu modificările temporale observate în efectul urmărit de la momentul T1 la T2.

3.4.3. Validitatea de construct

Constructele sunt esenţiale pentru funcţionarea oricărei ştiinţe prin cele trei roluri majore pe care le îndeplinesc (Shadish *et al.*, 200): (a) de a oferi o semnificaţie şi a facilita interpretarea rezultatelor unui studiu şi

integrarea acestora la nivel de teorii; (b) prin implicaţiile sociale, politice sau economice pe care le au, soldate deseori cu modificări terminologice, cum este cazul conceptelor cu potenţial de stigmatizare (termenul „handicapat” fiind înlocuit cu „persoană cu nevoi speciale” etc.); (c) de a facilita la avansarea cunoaşterii (sintagme precum „genomul uman”, „memorie procedurală” etc. erau necunoscute acum 100 de ani).

Caseta 3.2. *Exemplificarea utilităţii folosirii constructelor (o adaptare după Elmes, Kantowitz şi Roediger III, 2012)*

O serie de studii, atât experimentale, cât şi corelaţionale, din domeniul psihologiei transporturilor au evidenţiat mai multe rezultate disparate: (a) şoferii au nevoie de un timp mai mare de reacţie pentru a citi corect turometrul (turaţia motorului lor) pe măsură ce trebuie să vireze mai mult de volan (curbe mai abrupte), traficul este mai dens (sunt mai multe maşini în trafic), iar distanţa faţă de un vehicul periculos este mai mică (în faţa maşinii se află un tir imens, cu încărcătură periculoasă); (b) nu numai timpul necesar citirii turometrului este afectat de cei trei factori menţionaţi, se observă şi o scădere în viteza de manevrare a volanului (*steering wheel rate*); (c) mai mult, asemenea solicitări au efecte la nivel neurofiziologic, fiind observată o asociere între variabilitatea ritmului cardiac (scade variabilitatea) şi situaţiile de conducere descrise, cum ar fi conducerea maşinii în trafic dens sau în apropierea unor vehicule mari, ce induc un sentiment de nesiguranţă, respectiv în abordarea unor curbe periculoase.

Dacă ar fi să schiţăm logic relaţia dintre toate aceste variabile specificate am obţine o situaţie precum cea reprezentată în partea de sus a figurii 3.5. Pentru a face ordine în această reţea de legături şi pentru a simplifica interpretarea tuturor acestor relaţii, cercetătorii pot propune noi constructe care să ajute la o mai bună înţelegere (descriere) a situaţiei analizate şi care să ne permită să facem predicţii cu privire la evoluţia lucrurilor. În exemplul prezentat, un astfel de concept ar putea fi cel de nivel de solicitare (*workload*), care ţine seama de raportul dintre cerinţele mediului extern şi efectele asupra performanţei individului (partea de jos a figurii 3.5). Dacă nivelul de solicitare creşte, se poate ajunge la suprasolicitare, cu consecinţe negative asupra siguranţei în trafic. În plus, studii ulterioare pot extinde nivelul de generalizare (aplicabilitate) a unui asemenea construct la alte situaţii decât cele din trafic.

Acest tip de logică constituie baza oricărui demers din psihologie, încă de la începuturile sale ca ştiinţă (vezi teoriile motivaţionale bazate pe impulsuri şi nevoi pentru a explica diferite comportamente umane). Acest lucru se întâmplă pentru că în psihologie se

operează frecvent cu concepte abstracte (cum ar fi credinţele iraţionale, stima de sine, angajamentul organizaţional etc.) pentru a oferi un sens şi a ajuta la o mai bună înţelegere a diferitelor paternuri ale comportamentului uman. Spre deosebire însă de începuturile psihologiei, în prezent se acordă o mai mare importanţă cerinţei ca aceste noi constructe formulate în psihologie şi teoriile arondate acestora să fie conectate la rezultatele din alte ştiinţe (de exemplu, să fie sprijinite la un nivel implementaţional [biologic] de analiză). Bunăoară, teoria szondiană a genelor pulsiunilor şi a destinului nu este percepută ca fiind întemeiată pe constructe solide, deoarece prevederile ei sunt lipsite de o fundamentare biologică.

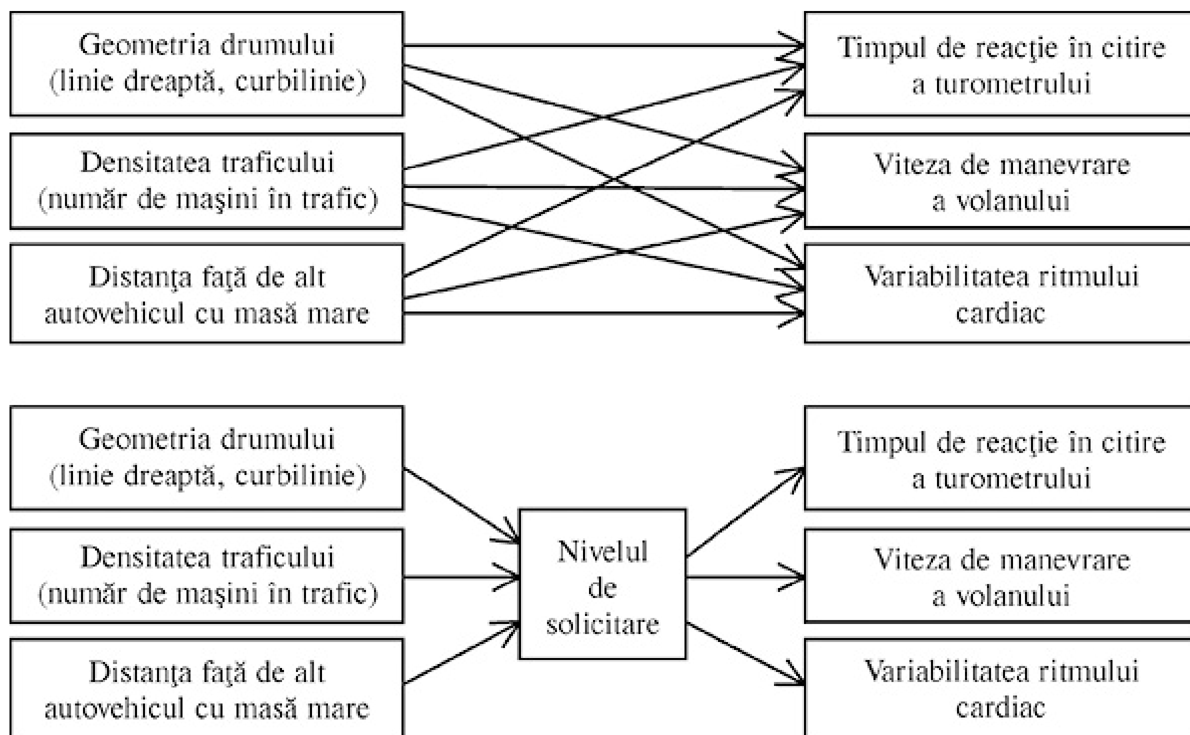


Figura 3.5. *Simplificarea înţelegerii şi interpretării realităţii prin introducerea de constructe*

Aşadar constructele psihologice sunt concepte abstracte inferate din comportamentele observate, pornind de la anumite modele teoretice de funcţionare a proceselor psihice considerate acceptabile la un moment dat (Kerlinger şi Lee, 2000). De exemplu, concepte precum inteligenţa generală, anxietatea, atribuirea sau nivelul de solicitare nu pot fi văzute, auzite sau atinse în realitate. Cu toate acestea, rolul lor este asemenea factorilor din analiza factorială: de a extrage aspecte comune din diferite

comportamente şi a oferi o interpretare pentru patternul observat (dacă un individ vorbeşte urât cu alţii, trânteşte uşi, îşi loveşte animalul de companie etc., spunem despre persoana respectivă că este *agresivă*).

Dată fiind importanţa constructelor pentru orice demers de cercetare, dar şi din cauza naturii lor abstracte, bazate pe interpretarea realităţii observate în baza unor modele ipotetice şi de explicare a minţii umane, adresarea aspectelor legate de validitatea de construct este sensibilă, dar esenţială pentru a interpreta corect rezultatele unui studiu. Pentru a înţelege mai bine distincţia dintre validitatea statistică, validitatea internă şi validitatea de construct, vom lua în considerare un posibil articol cu titlul: „Un studiu randomizat privind eficienţa psihoterapiei interpersonale la adolescenţii depresivi”. Validitatea statistică vizează măsura în care rezultatul statistic obţinut este unul corect (sunt evitate erorile de tip I sau II). Validitatea internă vizează măsura în care legătura observată poate fi privită în termeni cauzali (intervenţia respectivă este cauza pentru modificările observate la nivelul depresiei). Validitatea de construct vizează măsura în care cercetătorul a operaţionalizat corect studiul respectiv. În exemplul de mai sus, ar viza întrebări concrete precum: „suntem siguri că intervenţia oferită este corect denumită psihoterapie interpersonală?”; „suntem siguri că măsurătorile variabilei dependente constituie măsurători adecvate ale nivelului de depresie?”; „suntem siguri că participanţii la studiu sunt nominalizaţi corect ca fiind adolescenţi depresivi?”.

Pelham şi Blanton (2007) au sintetizat memorabil aceste caracteristici ale validităţii de construct spunând că are în vedere măsura în care cercetătorul a reuşit să facă (operaţionalizeze) ceea ce şi-a propus să facă, a spus că a făcut şi ar fi trebuit să facă (*how well was done what the researcher meant, said, or should have done*). În plus faţă de aceste

întrebări, tipice pentru orice tip de studiu empiric, în cazul studiilor care îşi propun un scop cauzal-explicativ (posibil numai prin studii randomizate), validitatea de construct poate contribui la un obiectiv suplimentar, acela de a găsi explicaţii pentru relaţia cauzală (de exemplu, „din ce motiv funcţionează psihoterapia interpersonală în combaterea depresiei?”; „care este explicaţia validă pentru acest rezultat?”). O altă frază memorabilă cu referire la validitatea de construct este oferită de Trochim (2006), parafrazând celebra afirmaţie din justiţie: „jur să spun adevărul, întregul adevăr şi numai adevărul”. Conform acestui autor, problema validităţii de construct se referă la măsura în care instrumentul utilizat reflectă „constructul, întregul construct şi numai constructul vizat”.

Însă, validitatea de construct nu se referă doar la corectitudinea măsurătorilor efectuate (de exemplu, măsurarea depresiei), aşa cum s-ar crede, ci vizează gradul de adecvare în modul de operaţionalizare şi implementare a întregului studiu, referindu-ne aici la participanţi, la contextul cercetării şi la variabilele măsurate în studiu, fie că este vorba de modul în care a fost implementată intervenţia (manipularea VI) sau de efectele observate (VD). Această viziune reprezintă o extindere făcută de Shadish *et al.* (2002), pornind de la modul iniţial de a conceptualiza validitatea de construct (Campbell şi Stanley, 1963; Cook şi Campbell, 1979) care se rezuma doar la nivelul de adecvare legat de operaţionalizarea variabilelor din studiu (manipularea experimentală, respectiv operaţionalizarea variabilei dependente sau a altor covariabile măsurate). Totuşi, deşi au trecut mai bine de zece ani de la această extindere a sferei validităţii de construct, numai trei dintre cele patru surse care pot viza validitatea de construct au cunoscut o dezvoltare continuă. Este vorba, în primul rând, de păstrarea în prim-plan a preocupărilor

pentru modul în care s-a conceput şi efectuat măsurarea variabilelor incluse în studiu (de exemplu, VI şi VD în studiile experimentale), respectiv de discuţia cu privire la validitatea de construct din perspectiva participanţilor incluşi în studiu. Luând în calcul această direcţie de dezvoltare, ne vom axa în cele ce urmează numai asupra acestor trei ipostaze importante din perspectiva validităţii de construct.

Nu în ultimul rând, o precizare importantă pentru studiile experimentale, evidenţiată de Susser, Schwartz, Morabia şi Bromet (2006), este aceea că atât validitatea internă, cât şi cea de construct vizează, în esenţă, identificarea variabilelor confundate. Totuşi, în cazul validităţii interne problema se pune în a identifica acele riscuri (variabile confundate) care ar fi condus la rezultate semnificative statistic în absenţa intervenţiei (acţiunii VI), fiind *extrinseci intervenţiei* (modului de manipulare a VI). Astfel, nu se poate distinge clar dacă intervenţia sau variabila confundată sunt cauza efectului observat. În schimb, în cazul validităţii de construct problema este identificarea acelor riscuri care au efect asupra variabilei dependente, dar care fac parte din intervenţie, fiind *intrinseci intervenţiei* (modului de manipulare a VI). În această situaţie nu se poate distinge clar sursa (explicaţia) relaţiei cauzale obţinute[11].

3.4.3.1. Validitate de construct cu implicaţii la nivelul participanţilor din studiu

Fie studiul non-experimental comparativ ce urmăreşte dacă există diferenţe între femeile cu rezidenţa în mediul rural şi cele cu rezidenţa în mediul urban în ceea ce priveşte frecvenţa utilizării prezervativelor. Problema cu acest tip de studiu, unde nu intervine niciun fel de manipulare experimentală (deci nu ridică probleme de validitate internă),

este legată de dificultăţile în operaţionalizarea participanţilor. Astfel, dacă nu sunt luate măsuri de control, este foarte posibil ca cele două loturi să difere şi cu privire la alte caracteristici studiate decât cea menţionată. De exemplu, este foarte posibil să existe o proporţie mai mare de femei cu studii superioare în lotul persoanelor din mediu urban decât al celor din mediul rural. De asemenea, s-ar putea să existe diferenţe între cele două loturi cu privire la modul în care se raportează comportamental la religie şi la preceptele religioase etc. De aceea, în lipsa unor măsuri elementare de control al participanţilor incluşi în studiu, am putea afirma la fel de bine că studiul urmăreşte diferenţele existente cu privire la utilizarea prezervativelor între femei cu studii superioare, care aderă la practicile religioase, şi cele fără studii superioare, care nu aderă la practici religioase. De asemenea, chiar dacă controlăm atent potenţiale variabile confundate, definitorii pentru participant, ar trebui acordată mai multă atenţie şi definirii a ceea ce înseamnă persoană cu rezidenţa în mediul rural, respectiv urban. De exemplu, poate că un procent de 25% dintre femeile incluse în lotul cu rezidenţa în mediul urban au mai puţin de un an vechime de când s-au mutat de la sat la oraş. Sunt în acest caz ele potrivite (reprezentative) pentru a fi incluse în grupul celor cu rezidenţă în mediu urban? Pentru a evita asemenea situaţii, cea mai bună soluţie ar fi aceea de a oferi definiţii operaţionale pentru ceea ce se înţelege prin femei cu rezidenţa în mediul urban, respectiv în mediul rural.

Problema validităţii de construct cu privire la participanţii incluşi în studiu este adresată de obicei în studiile randomizate controlate, precizându-se foarte clar ce fel de participanţi au fost eligibili pentru a fi incluşi în studiu şi care sunt criteriile de excludere. De exemplu, în studiul efectuat de David *et al.* (2008), care viza efectului terapiei cognitiv-comportamentale asupra adulţilor cu depresie majoră, se stipulează

explicit că au fost incluse doar persoanele care: (a) au fost diagnosticate cu depresie majoră după standardele DSM-IV; (b) au obţinut scoruri mai mari sau egale cu 20 la scala de depresie Beck; (c) au obţinut un rezultat de cel puţin 14 sau mai ridicat la scala de depresie Hamilton; (d) şi nu prezintă alte comorbidităţi psihiatrice după DSM-IV. De asemenea, au fost excluşi din studiu cei care urmau deja un tratament medicamentos sau psihoterapeutic, precum şi cei cu vârsta sub 18 ani sau cei de peste 70 de ani. Prin aceste precizări explicite se poate aprecia mult mai bine măsura în care modul de operaţionalizare asociat participanţilor corespunde constructului teoretic de adulţi suferinzi de depresie majoră.

3.4.3.2. Validitate de construct cu implicaţii la nivelul intervenţiei (manipulării VI)

În mod firesc, obţinerea unui rezultat semnificativ statistic în cadrul unui studiu experimental constituie un temei puternic pentru ca orice cercetător să considere că intervenţia testată este una eficientă sau că manipularea experimentală a produs efectul dorit. Validitatea de construct localizată la nivelul VI urmăreşte gradul de adevăr (încredere) în explicaţiile oferite cu privire la relaţia cauzală demonstrată. Ea nu pune în discuţie valoarea de adevăr cu privire la existenţa unei relaţii de tip cauză-efect (în studii experimentale) sau a unei relaţii de asociere (în studii non-experimentale), ci urmăreşte măsura în care explicaţiile psihologice oferite pentru aceste relaţii semnificative statistic au suport empiric.

De exemplu, în cazul intervenţiilor psihoterapeutice, Corday şi Pion (2006) subliniază faptul că deşi se demonstrează faptul că o intervenţie psihoterapeutică este eficientă (de exemplu, există o relaţie cauzală între expunerea la forma respectivă de psihoterapie şi ameliorarea

simptomatologiei depresive), efectul poate fi determinat de alte componente asociate intervenţiei decât cele utilizate ca explicaţie. Astfel, eficienţa unei intervenţii este dificil de explicat pentru că efectul poate fi cauzat de: (a) unele variabile confundate nespecifice, intrinseci intervenţiei (cum ar fi expectanţele cercetătorului, efectul placebo etc.); (2) alte explicaţii plauzibile care nu au fost verificate (explicaţii concurente); (3) unele cauze necunoscute în situaţia implementării problematice a intervenţiei (ceea ce s-a oferit clienţilor diferă semnificativ de ceea ce ar fi trebuit să li se ofere – aderenţa scăzută la prevederile manualului de intervenţie). De aceea, pentru a ne asigura că interpretarea unei relaţii cauzale este cea corectă, ar trebui să ne asigurăm că cele trei surse menţionate au fost controlate/nu constituie explicaţii plauzibile pentru rezultatul obţinut.

De aceea, vom discuta în paginile următoare principalele tipuri de probleme (riscuri) la adresa validităţii de construct cu referire specifică la cazul VI prezente în studii experimentale sau cvasiexperimentale. Acestea sunt reflectate sintetic în figura 3.6.

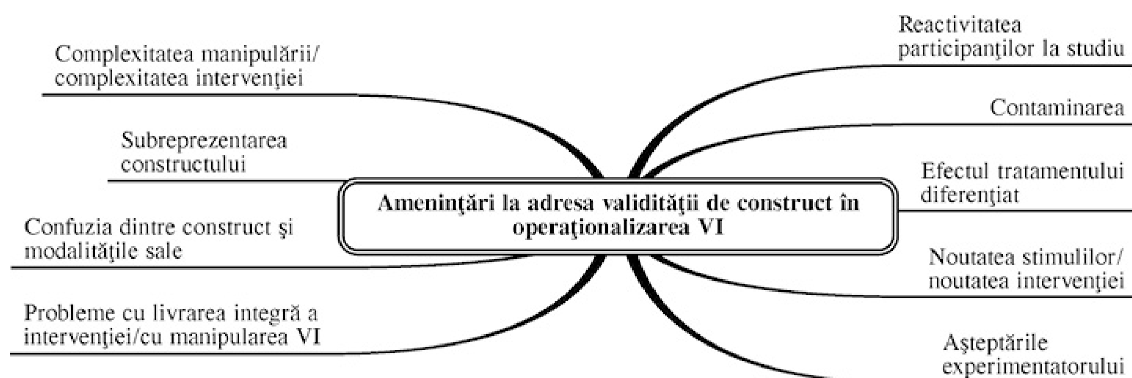


Figura 3.6. *Schiţa principalelor probleme*[12] *întâmpinate în adresarea validităţii de construct a variabilei independente*

3.4.3.2.1. Complexitatea manipulării VI/complexitatea intervenţiei

Printre cele mai subtile ameninţări la validitatea de construct (a modului în care a fost operaţionalizată variabila independentă) se regăseşte complexitatea constructului utilizat. Kazdin (2010) descrie succint cazul general al farmacoterapiei, în care grupul experimental (cel care primeşte medicamentul experimental) este comparat cu un grup de control (care nu primeşte nimic), iar rezultatul indică diferenţe semnificative statistic în favoarea grupului experimental. Un asemenea rezultat nu ridică probleme din perspectiva validităţii interne (presupunând că privitoare la aceasta au fost evitate). Problema apare atunci când cercetătorul îşi propune să explice rezultatul obţinut, cum ar fi ideea că medicamentul respectiv este eficient în reducerea simptomatologiei depresive, spre exemplu. Cele două grupe nu diferă numai la nivelul ingerării posibilei substanţe active. Simplu fapt că cei din grupul experimental primesc un tratament poate conduce la aşteptări pozitive din partea participanţilor incluşi în grupul experimental. Aceste aşteptări, cunoscute generic sub numele de efectul placebo, pot apărea chiar dacă medicamentul primit nu conţine nicio substanţă activă. Efectul apare pentru că pacienţii cred că medicamentul primit îi va face bine, le va ameliora problemele de sănătate. Rezultatul este uimitor! Studiile arată faptul că, deşi medicamentele antidepresive ajută la ameliorarea simptomatologiei/recuperare doar în aproximativ 50% din cazuri, un efect pozitiv se obţine şi în contextul în care se administrează placebo, pentru aproximativ 1 din 3 indivizi aflaţi, fără ştiinţa lor, în acest grup (Agenţia de Cercetare şi Politici de Sănătate Publică din SUA, 1999 *apud* Kazdin, 2010).

În cazul intervenţiilor psihologice, conceptualizarea ideii de placebo se concretizează adesea prin urmărirea unor factori nespecifici, cum ar fi

oferirea unor şedinţe de consiliere în care rolul consilierului este mai degrabă unul pasiv, de a oferi informaţii despre problema care necesită intervenţia şi de a asculta problemele avute de individul consiliat. Până la urmă, grupul experimental diferă de cel de control, nu numai prin forma de intervenţie primită (de exemplu, coaching cognitiv-comportamental), ci prin însăşi natura sarcinii. De exemplu, cei din grupul experimental participă regulat la şedinţe, spre deosebire de cei din grupul de control de tip „absenţa intervenţiei" şi primesc mai multă atenţie din partea echipei de cercetare. Participarea la asemenea şedinţe activează în mod diferit resursele psihologice ale indivizilor din grupul experimental faţă de cei din grupul de control, primii având o încredere mai mare că activităţile efectuate ar putea fi în beneficiul lor decât cei din grupul de control. Pentru a ne asigura că intervenţia respectivă şi nu efectul placebo este responsabil pentru rezultatul obţinut, ar fi dezirabil ca grupul de control să nu fie unul pasiv, ci să urmeze o intervenţie nespecifică pentru problema respectivă.

În sfârşit, chiar dacă efectul observat nu este cauzat de efectul placebo, rămâne să se stabilească ce anume din cadrul psihoterapiei a condus la efectul observat. Să fie adevărată explicaţia oferită de cercetător (de exemplu, reducerea cogniţiilor iraţionale autodepreciative reduce nivelul depresiei)? Sau ameliorarea este consecinţa altor factori psihoterapeutici precum încrederea ridicată în expertiza terapeutului, consolidarea alianţei terapeutice, modificări ale altor tipuri de cogniţii iraţionale decât cele autodepreciative etc.? Pentru a ne asigura de validitatea explicaţiei oferite este necesar apelul la studii care se focalizează pe identificarea unor mecanisme cauzale (vezi secţiunea 4.2.3), îndeosebi la studii care manipulează experimental gradientul (cantitatea de exerciţii) menit să conducă la relaţia cauzală observată.

Aceeaşi problemă a complexităţii VI din perspectiva modului în care a avut loc manipularea experimentală apare şi în alte tipuri de studii experimentale din psihologie, în afara celor care vizează testarea unor intervenţii cu impact aplicativ în domenii diverse (clinic, educaţional, organizaţional etc.). De exemplu, caseta 3.3 conţine descrierea unei asemenea situaţii în domeniul cercetării fundamentale din psihologia socială.

Aşadar, o consecinţă frecvent întâlnită în situaţiile care suferă de problema complexităţii constructelor este capacitatea redusă de a oferi explicaţii valide pentru rezultatele obţinute. Acestea rămân la stadiul unor discuţii speculative, al unor direcţii viitoare de cercetare.

Caseta 3.3. *Ilustrarea dificultăţilor de explicare a rezultatelor provenite din studii experimentale (cu validitate internă demonstrată; exemplul 2 – adaptare după Brewer, 2000)*

Exemplul ipotetic 1. *Ritmurile hip hop induc agresivitatea*

Este posibil ca, în cadrul unui studiu experimental, participanţii să fie supuşi într-una din cele două condiţii experimentale diferite: (a) să asculte albumele de muzică hip hop ale formaţiilor Paraziţii şi BUG Mafia; (b) să asculte şlagăre muzicale din anii ’80 din repertoriul Angelei Similea, respectiv al Corinei Chiriac.

Variabila dependentă este reprezentată de o măsură implicită a agresivităţii măsurată prin timpul de reacţie necesar completării unor anagrame (de exemplu, *lonag* rearanjat devine *golan*, lăsând la o parte discuţia cu privire la adecvarea acestei operaţionalizări a variabilei dependente din perspectiva validităţii de construct). Rezultatul studiului (fără probleme privind validitatea internă) sprijină existenţa unei relaţii cauzale, nivelul de agresivitate fiind mai ridicat în grupul celor care ascultă hip hop.

Explicaţia oferită de cercetător pentru acest rezultat vizează ca mecanism explicativ starea de spirit diferită transmisă de cele două genuri muzicale. O asemenea explicaţie este însă una pur speculativă, nefiind susţinută prin alte măsuri active în cadrul studiului. Acest lucru se întâmplă pentru că există o altă interpretare plauzibilă, aceea că diferenţa la nivel de agresivitate dintre cele două genuri muzicale este dictată de limbajul diferit (de texte), nu de ritmul pieselor. Atât timp cât cele două genuri muzicale diferă, atât la nivel de ritm, cât şi de „cuminţenie a textelor” niciuna dintre cele două interpretări nu depăşeşte un nivel speculativ, astfel că inferenţele explicative oferite nu au validitate de

construct. Desigur, un asemenea experiment ar putea fi îmbunătăţit printr-o simplă manipulare experimentală. În locul audierii unor formaţii româneşti (cu texte inteligibile pentru participanţii la studiu) se poate decide să se asculte aceleaşi genuri muzicale în interpretarea unor formaţii japoneze. În acest caz, diferenţele la nivel de text dispar, astfel că păstrarea unor diferenţe semnificative la nivelul agresivităţii nu ar mai putea fi explicate prin nivelul diferit de agresivitate prezent în conţinutul textelor. Însă problema interpretării ar fi rezolvată doar parţial, dacă alţi cercetători ar deschide o nouă linie de dezbare, exprimându-şi opinia că agresivitatea semnalată şi în muzica hip hop japoneză este transmisă nu prin ritmul melodic, ci prin diferenţe în tonalitatea vocii interpretului (mesajul paraverbal transmis).

Exemplul 2 (vezi Brewer, 2000; Wikipedia, 2012). *Disonanţa cognitivă creşte nivelul de atractivitate*

În studiul clasic realizat de Aronson şi Mills (1959) li se cere participanţilor la un experiment să discute despre psihologia sexului. Pentru a verifica dacă sunt pregătiţi să vorbească despre un asemenea subiect, participanţii au fost împărţiţi aleatoriu în trei grupe. Unii dintre participanţi au fost puşi iniţial să citească cu voce tare în faţa celorlalţi cuvinte cu conotaţie sexuală redusă (de exemplu, prostituată, virgină etc.) şi anumite pasaje de text cu acelaşi nivel moderat de conotaţii sexuale. Participanţii din al doilea grup au avut aceeaşi sarcină de făcut, doar că de această dată cuvintele şi pasajele citite aveau o conotaţie sexuală puternică (de exemplu, penis, vagin etc.). Subiecţii incluşi în cel de-al treilea grup, cel de control, nu au nicio sarcină de iniţiere în grup. Apoi toate grupurile au participat la o discuţie voit plictisitoare (direcţie indusă de către experimentator), cu tema: activitatea sexuală la animale. La finalul discuţiei de grup li s-a cerut tuturor să aprecieze cât de mult le-a plăcut grupul ad-hoc creat şi discuţia avută. Rezultatele obţinute au arătat că cei care au trecut prin condiţia experimentală cea mai dură (stânjeniţi de pronunţarea cu voce tare a unor cuvinte obscene în faţa unor necunoscuţi) au apreciat cel mai mult grupul de discuţii la care au participat.

Interpretarea autorilor pentru acest rezultat s-a făcut din perspectiva disonanţei cognitive. Astfel, plusul de apreciere în evaluarea grupului din care au făcut parte cei din condiţia experimentală cu conotaţii sexuale puternice a fost pus pe seama efortului suplimentar depus de ei pentru a depăşi starea de disconfort (jenă). Este posibil ca unul dintre gândurile pentru a-şi rezolva disonanţa legată de supliciul prin care au trecut să fi fost: „Până la urmă a meritat situaţia jenantă în care m-am aflat, pentru că grupul este chiar drăguţ”. Din nefericire pentru autori, această explicaţie a fost contestată de alţi cercetători, care au oferit o interpretare alternativă plauzibilă. Aceasta ar fi că nivelul ridicat de apreciere a grupului de discuţii din partea celor aflaţi în condiţia experimentală cu conotaţii sexuale puternice este un rezultat rezidual al excitării sexuale. Cu alte cuvinte, citirea pasajelor respective a condus la o creştere a excitării sexuale, iar acest efect s-a reportat şi în aprecierea calităţii grupului şi a discuţiei de grup.

Prima explicaţie oferită de autori nu se distinge de explicaţia alternativă din perspectiva suportului empiric, de aceea se poate afirma că interpretarea oferită este una speculativă, lipsită de validitate de construct.

Studii ulterioare au încercat să rezolve această dilemă (Gerard şi Mathewson, 1966 *apud* Brewer, 2000) prin eliminarea conotaţiilor sexuale din studiu, dar cu păstrarea unui context nefavorabil de iniţiere în discuţia de grup. Astfel, în locul citirii unor cuvinte cu conotaţie sexuală, li s-au aplicat participanţilor din condiţia experimentală adversivă nişte şocuri electrice, ca un test de sensibilitate emoţională. În locul discuţiei ulterioare despre sexul la animale, discuţia, derulată tot într-o manieră monotonă, a fost despre copiat şi alte forme de fraudă. Rezultatele obţinute din acest studiu au replicat conceptual ideea din cercetarea anterioară. Astfel, persoanele din condiţia experimentală care au fost supuse celui mai intens nivel de şoc electric au fost cele care au apreciat cel mai mult grupul şi discuţia de grup.

Observaţie! Printr-un asemenea demers se replică la nivel conceptual rezultatele studiului anterior şi se elimină varianta evaluării pozitive a grupului ca un efect al reportării excitării sexuale. Acest lucru nu înseamnă automat că prima explicaţie este confirmată. În acest caz, ca şi în cel al ştiinţei în general, ipotezele nu se confirmă, ci doar pot fi sprijinite până la proba contrarie. O astfel de probă contrarie pare fi cazul unui studiu efectuat recent pe porumbei (Singera şi Zentall, 2011), în care s-a arătat că porumbeii sunt mai încântaţi de recompense dacă acestea sunt primite după situaţii negative. Un astfel de rezultat ar elimina explicaţia bazată pe disonanţă cognitivă (dacă asumăm că porumbeii nu raţionalizează şi nu-şi justifică efortul!). În schimb, o altă explicaţie devine plauzibilă pentru toate cele trei studii descrise. Ea este bazată pe discrepanţa (distanţa) afectivă dintre starea iniţială percepută negativ şi starea finală pozitivă. În limbaj popular, bucuria pare a fi mai mare dacă vine după o stare de agonie (vezi sintagma „de la agonie la extaz"), decât dacă trecerea este de la o stare neutră sau uşor pozitivă la una de bucurie. Din acest punct de vedere, se poate spune că această explicaţie are un nivel ridicat al validităţii de construct deoarece: (a) reuşeşte să ofere o explicaţie plauzibilă pentru studiul efectuat pe porumbei şi să elimine totodată celelalte două variante explicative vehiculate; (b) să constituie explicaţii plauzibile în demersul de interpretare a studiilor anterioare. Atât studiul bazat pe aplicarea iniţială de şocuri electrice, cât şi cel bazat pe expunerea iniţială la un disconfort social pornesc de la o stare afectivă negativă, nu de la un punct de start neutru.

Însă, din punct de vedere al consolidării validităţii de construct, demersul de cercetare ar putea continua. De exemplu, ar fi utilă demonstrarea interpretării într-o manieră directă, cum ar fi cazul replicării celor două studii derulate pe oameni, însoţită de o verificare a manipulării experimentale. Spre exemplu, se poate demonstra că între condiţiile experimentale există diferenţe cu privire la intensitatea percepţiei negative a stării iniţiale, întrebând participanţii cât de bine se simt după sarcina de iniţiere (manipulare

confirmată dacă percepţia negativă este mai intensă în condiţia experimentală severă). Această dovadă ar constitui veriga lipsă necesară pentru interpretarea bazată pe discrepanţa afectivă dintre starea iniţială şi starea finală. Explicaţia şi-ar schimba statutul dintr-una plauzibilă, dar indirectă (rezultatele mai multor studii fiind interpretabile prin apelul la distanţa afectivă), într-o explicaţie „dovedită", sprijinită empiric printr-o serie de replicări conceptuale, cu un nivel ridicat al validităţii de construct.

3.4.3.2.2. Subreprezentarea constructului

O problemă frecvent întâlnită cu privire la validitatea de construct este aceea a operaţionalizării constructului teoretic printr-o singură variabilă sau printr-un set limitat de stimuli. Acest lucru duce la subreprezentarea constructului teoretic, problemă cunoscută în literatura de specialitate ca distorsiunea operaţionalizării singulare (*mono-operation bias*). Unii cercetători consacraţi în domeniul validităţii unei cercetări o consideră ca fiind cel mai serios risc privitor la validitatea de construct (Cook şi Campbell, 1979).

În sensul ei originar, această distorsiune are în vedere situaţia în care este utilizată o singură modalitate de reprezentare a unui construct, astfel încât nu se poate distinge dacă efectul observat este cauzat de constructul pe care îl reprezintă sau de aspectele particulare (specifice) ale stimulului (metodei respective) – vezi caseta 3.4. Ulterior, riscul şi-a extins sfera de cuprindere, fiind vizate orice situaţii în care constructul evaluat este operaţionalizat printr-un număr insuficient de stimuli sau de faţete ale constructului ori prin lipsa de reprezentativitate a acestora.

Caseta 3.4. *Ilustrarea subreprezentării constructului (distorsiunii mono-operaţionale)*

Exemplu ipotetic. *Impactul credibilităţii sursei asupra mesajelor cu rol persuasiv*
Într-un studiu experimental s-a vizat influenţa credibilităţii sursei asupra acceptării unui mesaj persuasiv luând cazul concret al Roşiei Montane şi al beneficiilor economice rezultate în urma începerii exploatării zăcământului aurifer. În cele două grupe experimentale utilizate a fost transmis acelaşi mesaj (pentru a controla impactul

conţinutului acestuia), fiind însă manipulată sursa (personajul) care se presupune că a formulat mesajul respectiv.

Primului grup experimental i s-a transmis că este vorba despre o femeie casnică, în vârstă de 45 de ani, mamă a trei copii, care se ocupă în timpul liber de croşetat pentru a face un ban în plus. Aceasta a lucrat o vreme la o întreprindere aflată la 20 de km distanţă, dar apoi a renunţat, preferând să se ocupe de treburile gospodăreşti şi de educaţia copiilor, bazându-se financiar doar pe ajutorul social de la stat şi alocaţia copiilor. Celui de-al doilea grup i s-a specificat că mesajul este transmis de un analist financiar, doctor în ştiinţe economice cu o temă privind rentabilitatea economică a sectorului minier, bursier Fullbright, având numeroase publicaţii de specialitate privind dezvoltarea economică a regiunilor miniere defavorizate, dar şi o capacitate bună de a comunica mesaje accesibile publicului larg pe înţelesul acestuia. Cele două situaţii experimentale diferă, aşadar, prin nivelul diferit de expertiză a personajului despre care se afirmă că este *sursa* textului persuasiv prezentat.

Problema cu acest gen de studii este faptul că scenariile prezentate reprezintă cazuri particulare singulare pentru constructele uzitate, fapt ce poate conduce la dificultăţi în explicarea clară a motivelor pentru care s-au înregistrat diferenţe între grupe. De exemplu, categoria generală a experţilor economici vizează o multitudine de exemplare diferite. De aceea, este posibil ca diferenţele observate să fi fost altele dacă în descrierea celui de-al doilea personaj, înlocuiam sintagma bursier Fullbright, cu aceea de angajat al Gabriel Gold Corporation (compania canadiană cu interese economice directe în exploatarea aurului din zonă). De asemenea, dacă studiul ar fi fost efectuat pe studenţi de la Litere, iar în locul mamei ar fi fost prezentată o tânără absolventă de limbi străine, reîntoarsă pe meleagurile natale, impactul ar fi putut fi altul.

Soluţia necesară pentru a evita asemenea situaţii este aceea a apelului la un set de cazuri diferite pentru a descrie cele două categorii avute în vedere: fără expertiză economică, respectiv cu o asemenea expertiză. Bunăoară, pentru prima condiţie experimentală s-ar apela, în locul unei singure descrieri, la cinci personaje diferite (drept caracteristici demografice, stil de viaţă), dar care împărtăşesc acelaşi atribut (lipsa expertizei în domeniul economic). Acelaşi demers s-ar putea realiza pentru cel de-al doilea grup experimental, apelându-se la cinci personaje diferite, toate cu expertiză în domeniul economic. Fiecare participant în acest studiu modificat, ar fi în continuare expus unui singur personaj (sursă), selectat în mod aleatoriu dintre personajele dezvoltate pentru condiţia experimentală de care aparţine. Dacă rezultatele indică diferenţe între cele două grupe experimentale în direcţia aşteptată, vom avea o mai mare de încredere în a afirma că nivelul de expertiză al sursei influenţează actul persuadării dincolo de conţinutul mesajului transmis. În absenţa acestor exemplare multiple reprezentative pentru acelaşi construct, diferenţele dintre condiţiile experimentale ar putea fi cauzate de alte aspecte specifice care diferenţiază cazurile

unice comparate (primul personaj fiind o femeie casnică, cel de-al doilea un bărbat activ în câmpul muncii etc.).

O problemă similară cu cea precizată în caseta 3.4 poate apărea în cazul studiilor randomizate controlate care testează eficienţa unei intervenţii. Kazdin (2010) surprinde corect această problemă. De exemplu, fie cazul în care se urmăreşte compararea eficienţei intervenţiei A cu cea a intervenţiei B, intervenţia A fiind implementată de doi terapeuţi specialişti în intervenţia A, în timp ce intervenţia B este livrată de un singur specialist. Rezultatul obţinut sprijină ipoteza că intervenţia B este mai eficientă decât intervenţia A în diminuarea problemei semnalate. Deşi din perspectivă statistică şi a validităţii interne, afirmaţia poate fi validă, există o explicaţie alternativă pentru rezultatul obţinut. Este posibil ca diferenţa observată să apară pentru că terapeutul specialist în intervenţia B este un profesionist mai bun decât cei care au implementat intervenţia A. Mai mult, chiar dacă ambele tipuri de intervenţie ar fi implementate de acelaşi psihoterapeut (care deţine acreditări atât pentru intervenţia A, cât şi pentru B), este posibil ca acesta să se simtă mai confortabil şi mai eficient într-o anumită formă de psihoterapie.

Din cauza acestor limite interpretative, se recomandă ca atunci când se doreşte compararea eficienţei unor forme diferite de intervenţie să se apeleze la un număr mai mare de persoane care să implementeze fiecare condiţie experimentală, pentru ca la final să se analizeze măsura în care compararea eficienţei intervenţiilor este moderată de eventuale diferenţe surprinse în calitatea terapeuţilor care au livrat intervenţiile respective.

3.4.3.2.3. Confuzia dintre construct şi modalităţile sale

Să ne imaginăm că este testată eficienţa unui program de coaching executiv menit să reducă fluctuaţia de personal. Durata întregului program este de 10 ore cumulate cu ocazia a cinci întâlniri faţă în faţă între consultant şi manager. Rezultatul obţinut nu sprijină ideea eficienţei acestui program (diferenţele dintre pretest şi posttest fiind nesemnificative statistic). Însă interpretarea corectă în acest caz ar fi una uşor diferită. Nu putem afirma că programul de coaching în ansamblul său este unul ineficient, ci că programul de coaching în formatul scurt cu cinci şedinţe (totalizând 10 ore de coaching) este unul ineficient. Acest lucru este cu atât mai probabil cu cât un program de coaching mai lung, de exemplu, totalizând 50 de ore, să aibă un nivel de eficienţă demonstrat empiric.

Acest gen de probleme este frecvent întâlnit în domeniul psihologiei experimentale, în care o importanţă deosebită revine calibrării stimulilor. De exemplu, un efect de amorsare semantică (*conceptual priming*) poate fi constatat dacă stimulul amorsă este prezentat timp de 200 ms înainte de apariţia stimulului-ţintă, în timp ce efectul de amorsare este absent dacă stimulul-amorsă este prezentat doar 150 de ms înainte de apariţia stimulului-ţintă. În acest context nu spunem despre efectul de amorsare că nu funcţionează, ci că efectul nu apare dacă expunerea la stimulul-amorsă este mai mică de 200 de ms. Un exemplu similar, dar mai facil de înţeles vine din sfera medicinei. Putem spune despre un anumit antibiotic că este eficient în combaterea unei maladii doar dacă discutăm la un nivel generic. În realitate, antibioticul respectiv este eficient în lupta cu maladia respectivă doar dacă este luat în anumite doze prestabilite experimental.

3.4.3.2.4. Probleme cu livrarea integră a intervenţiei/acurateţea manipulării VI

Între ceea ce este planificat pe hârtie şi ceea ce se întâmplă în realitate pot exista uneori diferenţe importante ca urmare a devierii de la protocolul (manualul) intervenţiei, de la procedura standard stabilită. Conform lui Carroll *et al.* (2000), orice intervenţie oferită ar trebui descrisă în termeni de: (a) doză oferită – frecvenţa, numărul şi durata şedinţelor; (b) ingrediente active ale intervenţiei – specificul intervenţiei şi premiza generală urmărită pentru a produce modificările dorite; (c) condiţii în care sunt administrate; (d) evaluarea măsurii în care intervenţia a fost livrată adecvat. De exemplu, în cazul ipotetic al unui studiu ce ar viza efectul terapiei raţional-emotive în direcţia reducerii depresiei comparativ cu un grup de control de tipul listei de aşteptare, o posibilă descriere a intervenţiei oferită grupului experimental ar putea fi:

> Terapia raţional-emotivă implementată în grupul experimental: (a) s-a derulat bisăptămânal, timp de 20 de şedinţe cu durata de 1 oră; (b) în cadrul acestora s-au urmărit identificarea, disputarea şi înlocuirea credinţelor iraţionale ale participanţilor cu altele raţionale, îndeosebi a celor care conţin referiri autodepreciative (*self-downing irrational beliefs*), considerate a fi responsabile pentru apariţia/menţinerea depresiei; (c) şedinţele s-au derulat individual, prin intermediul a patru terapeuţi certificaţi REBT, cu o experienţă clinică medie de 5 ani de zile; (d) intervenţiile s-au derulat conform specificaţiilor prezentate în manualul intervenţiei, aderenţa la tratament fiind verificată prin sondaj cu ajutorul înregistrărilor audio existente.

Cu alte cuvinte putem spune despre o intervenţie că are probleme cu livrarea integră (implementarea fidelă) dacă: (a) nu dispune de un manual de intervenţie care să ofere recomandări standard cu privire la ceea ce este de făcut în cadrul intervenţiei; (b) dispune de un asemenea manual, dar din verificările efectuate se constată o aderenţă scăzută a terapeuţilor/trainerilor la specificaţiile din manual (fie din cauza unui nivel precar de pregătire al celor care implementează intervenţia, fie din cauze motivaţionale diverse, fie de la un stil de lucru orientat mai degrabă

spre nevoile clientului decât spre atingerea obiectivelor planificate). În sfârşit, o a treia cauză posibilă pentru apariţia unei discrepanţe între ceea ce se intenţionează a se livra şi ceea ce se întâmplă în realitate poate fi (c) un nivel scăzut al complianţei clienţilor la cerinţele formulate de către terapeuţi/traineri.

Din motivele invocate mai sus, se obişnuieşte ca, în studiile randomizate controlate care urmăresc eficienţa unei intervenţii, să fie colectate date suplimentare cu privire la integritatea intervenţiei livrate, aşa cum a fost ea planificată iniţial, cu referire specifică la punctele (b) şi (c) menţionate în paragraful anterior. Astfel, se vorbeşte de aderenţa terapeuţilor la tratament dacă, în urma examinării unui eşantion de şedinţe din totalul şedinţelor de terapie înregistrate, se constată o corespondenţă între ceea ce era specificat în manualul intervenţiei cu privire la şedinţa respectivă şi ceea ce s-a întâmplat de fapt. Corespondenţa nu vizează aspecte de detaliu (replici etc.), ci vizează obiectivul şedinţei şi repertoriul de exerciţii/sarcini de efectuat în vederea atingerii obiectivului respectiv. Tot prin intermediul unor aprecieri, mai degrabă cu caracter subiectiv, obţinute prin intermediul unui expert neimplicat în studiu, se poate evalua competenţa celor care livrează intervenţia. În sfârşit, se poate de asemenea evalua nivelul de complianţă al participanţilor la intervenţia primită (măsura în care acceptă serviciile primite şi răspund pozitiv solicitărilor formulate de terapeuţi – de exemplu, prin efectuarea temelor pentru acasă).

Aspecte legate de integritatea livrării (manipulării) variabilei independente apar şi în alte contexte experimentale decât cele care vizează testarea eficienţei unei intervenţii. În acest context al discuţiei pot fi evidenţiate două nuanţe ale integrităţii cu referirea la adecvarea manipulării experimentale. Pe de o parte, transpare ideea implementării

standardizate a manipulării VI (absenţa unor artefacte introduse de erori experimentale). Un astfel de exemplu este cel determinat de respectarea protocolului standardizat în activarea condiţiilor experimentale şi a altor aspecte procedurale implicate (de exemplu, dacă studiul solicită ca subiectul să fie poziţionat la 30 cm de un monitor cu diagonală de 17 inch, atunci ar trebui să se respecte această cerinţă; dacă studiul presupune un anumit instructaj standard, atunci acesta ar trebui livrat integral şi celor care dau impresia că au înţeles repede sarcina pe care o au de efectuat etc.). Pe de altă parte, integritatea tratamentului în contexte experimentale, cum sunt cele de laborator, face referire directă la conceptul cunoscut sub denumirea de verificarea reuşitei manipulării experimentale (*manipulation check*) – vezi caseta 3.5 pentru ilustrarea conceptului şi importanţa acestuia pentru validitatea de construct a studiului.

Caseta 3.5. *Verificarea reuşitei manipulării experimentale*

Exemplu ipotetic. *Dispoziţia afectivă negativă scade performanţa la probe de viteză mentală*

Pentru a verifica validitatea acestei afirmaţii, s-a apelat la un experiment în care participanţii au fost împărţiţi în trei condiţii experimentale. Astfel, în prima fază a experimentului unii au citit o istorioară amuzantă cu un final fericit pentru personajul principal al poveştii, alţii au citit o istorioară tristă cu un final negativ de poveste, iar un al treilea grup a citit un text abstract, neutru din perspectiva mesajului şi a deznodământului oferit. Această manipulare experimentală a fost implementată pornind de la premisa că celor din prima condiţie experimentală povestirea amuzantă le va induce o stare de spirit optimistă, pozitivă, celor din a doua condiţie experimentală povestirea tristă le va induce o dispoziţie afectivă negativă, iar cei din cea de-a treia condiţie experimentală vor avea o dispoziţie afectivă neutră. În a doua fază a experimentului, toţi participanţii la studiu, indiferent de condiţia lor experimentală au completat un test de viteză perceptivă, luându-se în calcul performanţa lor ca număr de elemente parcurse corect într-un interval de 5 minute. Rezultatele statistice au arătat că există diferenţe între cele trei grupe, cei din condiţia experimentală 2 (povestire cu final trist) având cel mai scăzut nivel al performanţei la sarcina de viteză.

Pentru a ne putea asigura că inferenţa de la baza acestui studiu este una corectă (că manipularea experimentală este adecvată) este necesar să arătăm mai întâi că există diferenţe între cele trei grupe comparate cu privire la nivelul dispoziţiei lor afective, nu doar să presupunem acest lucru. De exemplu, subiecţii ar putea răspunde după finalizarea istorioarei la un set scurt de itemi care să reflecte starea lor de moment (cât de bine sau cât de trişti se simt în momentul respectiv?), iar apoi să completeze proba de viteză. Vom considera că manipularea experimentală a reuşit dacă între cele trei grupe se constată diferenţe semnificative cu privire la dispoziţia lor afectivă din momentul respectiv, în direcţia anticipată încă din momentul planificării studiului (de exemplu, că dispoziţia afectivă negativă are cel mai înalt nivel în condiţia experimentală în care s-a lecturat povestea cu un final trist, diferenţa faţă de celelalte două grupe fiind una semnificativă statistic).
Deşi poate părea redundant, un asemenea demers constituie o măsură necesară pentru a ne asigura că *ceea ce s-a intenţionat* (obţinerea unor variaţii în dispoziţia afectivă a participanţilor în funcţie de condiţia lor experimentală) *s-a petrecut în realitate*. Pot exista situaţii când manipularea experimentală eşuează, din diverse motive precum: (a) variaţii prea mici în intensitatea stimulilor (de exemplu, deşi alcoolul scade timpul mediu de reacţie, se poate ca acest lucru să nu fie observat dacă se compară două loturi care consumă cantităţi reduse de alcool (100 ml de bere *vs* 200 ml de bere); (b) caracterul artificial (nivelul redus de realism al condiţiei experimentale), fapt ce îi poate determina pe participanţi să răspundă diferit de comportamentul lor natural.
În sfârşit, verificarea manipulării experimentale poate juca un rol important în eliminarea unor explicaţii alternative. De exemplu, în cazul descris mai sus o altă explicaţie posibilă pentru scăderea vitezei de percepţie are legătură mai degrabă cu alterarea percepţiei timpului (perioada de timp petrecută pentru lecturarea povestirii triste pare a fi mai lungă), iar acest efect poate fi reportat în proba de viteză, participanţii având impresia că au mai mult timp la dispoziţie pentru completarea ei, fapt pentru care adoptă un ritm ceva mai puţin alert de lucru.
Pentru a elimina o asemenea explicaţie alternativă, la finalul povestirii participanţii pot fi rugaţi atât să estimeze timpul care a trecut de la momentul începerii lecturării povestirii, cât şi să se autoevalueze cu privire la dispoziţia lor afectivă curentă. Dacă nu se constată diferenţe semnificative între grupe cu privire la durata timpului estimat, dar apar diferenţe între grupe în direcţia aşteptată cu privire la starea lor afectivă, putem spune că explicaţia bazată pe diferenţe în dispoziţia afectivă este validă. Din acest punct de vedere, verificarea reuşitei manipulării experimentale poate oferi explicaţii utile, verificabile ulterior prin studii ce vizează identificarea mecanismelor schimbării responsabile pentru apariţia unei relaţii de tip cauză-efect.

În urma introducerii verificării reuşitei manipulării experimentale şi a măsurării existenţei unei relaţii dintre VI şi VD rezultă patru situaţii posibile (Kazdin, 2010): (a) efectele apar atât la nivelul manipulării experimentale, cât şi la nivelul variabilei dependente; (b) manipularea experimentală pare a fi reuşit (există diferenţe între grupe), dar nu există diferenţe la nivelul VD; (c) manipularea experimentală pare a nu fi reuşit (nu există diferenţe între grupe), dar acestea apar la nivelul VD; (d) nici cu ocazia verificării reuşitei manipulării experimentale, nici la nivelul diferenţelor dintre grupe cu privire la VD, nu se obţin rezultate semnificative statistic. Interpretarea semnificaţiilor posibile pentru aceste rezultate este reflectată în tabelul de mai jos.

Tabelul 3.8. *Interpretarea rezultatelor ca urmare a dinamicii dintre efectul la nivelul VD şi reuşita (nereuşita) manipulării experimentale*

Cazul A	*Cazul B*
Manipularea VI reuşită ($p < 0,05$), efect VD prezent ($p < 0,05$) VI are efect asupra VD (există o relaţie de tip cauză-efect).	Manipularea VI reuşită ($p < 0,05$), efect VD absent ($p > 0,05$) (i) VI nu are efect asupra VD (nu există o relaţie cauzală) sau (ii) deşi manipularea VI a reuşit, sunt probleme de măsurare a VD care necesită o amplificare a manipulării VI. *Exemplu (ii): manipularea VI a reuşit să modifice cogniţii/atitudini, dar nu schimbă comportamente (VD reflectând în acest caz comportamente). Pentru ca efectele să apară şi la nivel comportamental, ar fi necesar să se intensifice manipularea VI (de exemplu, în locul unui program de dezvoltare personală cu durata de 20 de ore să se aplice unul de 100 de ore).*
Cazul C Manipularea VI nereuşită ($p > 0,05$), efect VD prezent ($p <$	*Cazul D* Manipularea VI nereuşită ($p > 0,05$), efect VD absent ($p > 0,05$)

0,05) (i) Diferenţele obţinute la nivelul VD sunt întâmplătoare (eroare de tip I) sau sunt cauzate de un alt factor decât VI. (ii) Verificarea reuşitei manipulării experimentale s-a făcut greşit (probleme psihometrice). *Exemplu (ii): s-a reflectat un alt construct decât VI; s-au formulat ambiguu/neclar itemii din procedura de verificare a reuşitei manipulării, fapt care a afectat fidelitatea acesteia.*	Nu este clar dacă VI are efect asupra VD. Manipularea a fost prea slabă pentru a produce efecte, prin urmare ipoteza studiului nu a fost realmente testată. Sunt necesare modificări în modul de manipulare a VI.

3.4.3.2.5. Reactivitatea participanţilor la studiu

Participanţii la studiile psihologice, cu atât mai mult la cele derulate în condiţii de laborator, au tendinţa de a reacţiona diferit de felul lor natural de a fi, atunci când ştiu că sunt supuşi unei evaluări psihologice (vezi caseta 3.6). Acest lucru constituie un risc privitor la validitatea de construct deoarece ar fi dificil de precizat dacă explicaţia pentru efectul obţinut se regăseşte în caracteristicile intervenţiei sau este doar o consecinţă a reactivităţii participanţilor.

Caseta 3.6. *Ilustrarea reactivităţii participanţilor ca risc privitor la validitatea de construct*

Reactivitatea participanţilor

Exemplificare: Într-un studiu cvasiexperimental se urmăreşte dacă un material persuasiv despre un personaj politic modifică atitudinea potenţialilor alegători faţă de acesta. Astfel, atât înainte, cât şi după prezentarea unui film documentar de 20 de minute, participanţii sunt rugaţi să evalueze gradul de simpatie pe care îl au faţă de personajul

respectiv, pe o scală de la 1 [extrem de antipatic] la 9 [extrem de simpatic]. Conţinutul filmului documentar era unul negativ la adresa acestuia. Rezultatul obţinut sprijină ideea eficienţei unor astfel de materiale, nivelul de simpatie pentru personajul evaluat fiind mai redus la posttest decât la pretest.

Prin reactivitatea participanţilor[13] se înţelege, la nivel general, ideea că participanţii dintr-o cercetare sunt conştienţi de acest rol, fapt pentru care se pot comporta diferit faţă de felul lor natural sau pot decide să ofere răspunsuri consistente cu ipotezele implicite pe care cred că le testează studiul. De exemplu, în cazul descris, este foarte probabil ca mulţi dintre participanţi să intuiască ce se urmăreşte prin studiul respectiv, cum ar fi „în ce măsură pot fi ei influenţaţi de materiale cu rol propagandistic?". În baza acestei intuiţii cu privire la scopul studiului, ei pot decide să ofere răspunsuri care să-l mulţumească pe experimentator, în direcţia reducerii nivelului de simpatie pentru personaj. Alţii pot adopta poziţii opuse, de rezistenţă faţă de materialul persuasiv. Această situaţie este cunoscută în literatura de specialitate sub denumirea de caracteristicile sarcinii (*demand characteristics*) (Orne, 1962 *apud* Kazdin, 2010). Cel mai adesea, consecinţa acestui efect este de a sprijini ipoteza intuită, fiind mai mulţi participanţi complianţi (sau preocupaţi să lase o impresie bună), decât participanţi rezistenţi (care „nu se lasă păcăliţi/manipulaţi" de către experimentator).

Soluţii uzuale: Chestionarea participanţilor la finalul experimentului (debriefing)[14] cu privire la ce cred că s-a urmărit prin studiu (care erau ipotezele vizate) şi analiza statistică a rezultatelor separat la cei care au intuit, respectiv nu au intuit corect despre ce a fost vorba în studiu. Reactivitatea este prezentă dacă eficienţa intervenţiei (manipulării experimentale) apare numai la cei care au intuit ce se urmărea prin studiul respectiv/este mai intensă în cazul acestora.

3.4.3.2.6. Noutatea intervenţiei

Acest tip de risc reprezintă un tip particular de reactivitate a participanţilor. Orice intervenţie nouă, o inovaţie într-un domeniu, poate contribui la creşterea entuziasmului, a energiei şi aşteptărilor cu privire la utilitatea unei asemenea intervenţii (pe principiul că dacă este un produs sau serviciu de ultimă oră ar fi mai bun decât altele existente pe piaţă de ceva timp). În cazul unor rezultate semnificative statistic este greu de stabilit dacă acest fapt se datorează eficienţei intervenţiei în sine sau este

o consecinţă a percepţiei pozitive (a etichetelor de tip „nou”, „revoluţionar”, „de ultimă generaţie”).

Un tip aparte de efect, ca rezultat al combinaţiei dintre noutate şi reactivitatea participanţilor este efectul Hawthorne (Roethlisberger şi Dickson, 1939 *apud* Shadish *et al.*, 2002). Acesta se referă la situaţiile în care apar modificări semnificative între momentul dinainte şi cel de după introducerea unei intervenţii, dar acest lucru nu este generat de eficienţa intervenţiei, ci de situaţia experimentală în sine (participanţii sunt conştienţi de statutul lor social special ca subiecţi în cercetare, se bucură de o atenţie specială şi sunt beneficiarii unor intervenţii „noi”, diferite de activităţile de uzură în care au fost implicaţi până atunci).

Pentru a diferenţia dacă efectul observat este datorat intervenţiei în sine sau elementului de noutate, se poate urmări dacă intervenţia îşi păstrează aceeaşi magnitudine a mărimii efectului (un nivel similar de eficienţă), pe măsura trecerii timpului.

3.4.3.2.7. Contaminarea

Prin contaminare se are în vedere situaţia când o parte sau întreaga intervenţie destinată grupului experimental devine cunoscută şi grupului de control. De exemplu, un program de consiliere menit să îmbunătăţească comunicarea asertivă este implementat într-o clasă de elevi (a IX-a A), în timp ce elevii din clasa a IX-a B sunt incluşi în studiu drept grup de control. În pretest, cele două clase au înregistrat niveluri similare în abilitatea de a comunica asertiv. Totuşi, după fiecare şedinţă de consiliere, câţiva elevi din clasa a IX-a A se întâlnesc cu prietenii lor din clasa a IX-a B, cărora le povestesc din secretele deprinse în orele de consiliere. Aceştia devin interesaţi de acest subiect şi decid să exerseze şi ei sarcinile primite ca temă pentru acasă de către cei din grupul

experimental. Ca urmare a acestui fapt, se obţine un rezultat nesemnificativ statistic, astfel că eficienţa programului de consiliere nu este sprijinită de datele empirice colectate. Contaminarea, deşi este inclusă clasic în rândul riscurilor privitoare la validitatea de construct, poate fi percepută ca un pericol pentru validitatea internă, deoarece creşte probabilitatea apariţiei unei erori de tip II (să nu fie identificate diferenţe semnificative statistic, când ar fi trebuit să apară, dar acestea au fost estompate prin faptul că şi cei din grupul de control au beneficiat parţial sau în întregime de facilităţile oferite celor din grupul experimental). Astfel, în funcţie de gravitate (nivelul de extindere al contaminării), consecinţa poate consta în diminuarea intensităţii efectului produs de intervenţie (diminuarea mărimii efectului) sau chiar în anularea efectului produs de intervenţie (diferenţe nesemnificative între cele două grupe).

O soluţie posibilă în contextele cu risc crescut de apariţie a contaminării (de exemplu, în cazul în care grupele comparate sunt constituite din două clase şcolare paralele din cadrul aceleiaşi şcoli) este aceea de a apela la studii experimentale cu randomizare la nivel de cluster.

3.4.3.2.8. Efectul tratamentului diferenţiat

Acest efect se manifestă la participanţii din grupul de control atunci când: (a) aceştia fac parte dintr-un grup de control pasiv de tipul absenţei intervenţiei sau a listei de aşteptare (detalii în secţiunea 4.1.4); (b) în cazul în care deşi beneficiază de o intervenţie activă, ca membri ai grupului de control, ei ştiu că serviciul pe care-l primesc nu este cel inovativ, experimental (potenţial mai bun decât celelalte forme de intervenţie existente la vremea respectivă). Într-un astfel de caz este foarte posibil ca rezultatele grupului de control, în consecinţă întregul demers comparativ,

să fie afectate: (a) fie prin demoralizarea membrilor din grupul de control pentru faptul că nu beneficiază de facilităţile oferite celor din grupul experimental; (b) fie prin comportamente de compensare (pentru a nu rămâne în urmă faţă de cei din grupul experimental). În plus, în cazul grupelor de control pasiv cercetătorii pot interveni oferindu-le beneficii compensatorii pentru neincluderea în condiţia experimentală activă, pentru a-i motiva să rămână participanţi în studiu. Această atenţie sporită şi eventualele beneficii compensatorii oferite, s-ar putea să producă un efect activ nespecific, cauzat de simpla atenţie/grijă acordată participanţilor din grupul de control.

Consecinţele tratamentului diferenţial variază în funcţie de predominanţa efectului observat. Astfel, dacă în grupul de control apare mai frecvent demoralizarea membrilor, se măreşte diferenţa dintre condiţiile experimentale (nu mai ştim cu certitudine dacă diferenţele observate sunt cauzate de intervenţie, ca urmare a nivelului motivaţional scăzut din grupul de control sau a diferenţelor cu privire la expectanţa succesului între condiţiile experimentale). În schimb, dacă membrii din grupul de control adoptă mai degrabă comportamente compensatorii sau dacă sunt beneficiarii unui surplus de atenţie şi/sau măsuri compensatorii, este posibil ca distanţa dintre grupele experimentale să se reducă. Astfel că, în situaţia unui rezultat nesemnificativ statistic, nu vom cunoaşte dacă rezultatul este cauzat de ineficienţa intervenţiei sau de tratamentul diferenţiat cu rol compensator pe care îl adoptă sau de care se bucură participanţii incluşi în grupul de control.

Cea mai eficientă şi întâlnită soluţie pentru acest tip de risc este oferirea unor intervenţii nespecifice de tip placebo participanţilor din grupul de control sau mascarea rezultatului distribuirii randomizate (designuri de tip orb) acolo unde acest lucru este posibil.

3.4.3.2.9. Aşteptările experimentatorului

Cele mai cunoscute efecte din această categorie sunt efectele de autoîndeplinire a profeţiei, respectiv Pygmalion, care apar drept consecinţe ale unor aşteptări diferenţiate (Rosenthal, 1966 *apud* Kazdin, 2010). În zona oricărei forme de intervenţie, ele se pot observa sub forma unor schimbări în tonul vocii, în postură, în expresia facială, în tipul de întrebări formulate sau în modul de adresabilitate faţă de client. Dacă acesta face parte din condiţia experimentală activă (comparativ cu un grup de control pasiv) sau din condiţia experimentală preferată de către experimentator sau despre care acesta are impresia că este mai eficientă decât alte forme de intervenţie (în comparaţii cu un grup de control activ, perceput mai puţin pozitiv), este posibil ca diferenţele observate la finalul intervenţiei să nu fie cauzate de diferenţe în eficienţa acestora, ci de diferenţe în nivelul de aşteptare, în modul de implicare, în modul de relaţionare cu clienţii aparţinând grupelor diferite. Desigur, acest tip de efect poate constitui un risc doar în contextul în care diferenţa observată apare în favoarea grupului preferat.

Între procedurile care pot fi implementate pentru a reduce riscul apariţiei acestui tip de ameninţare se regăsesc (Shadish *et al.*, 2002): (a) utilizarea mascării condiţiei din care fac parte indivizii evaluaţi şi/sau a ipotezelor urmărite în studiu (de exemplu, prin implementarea unor designuri de tip orb pentru evaluatori şi/sau cei care implementează intervenţia); (b) reducerea la minimum a contactului dintre examinator şi participanţii la studiu. Alte măsuri care pot fi luate sunt: (c) utilizarea unui număr mai mare de asistenţi de cercetare care implementează studiul, pentru a putea identifica un eventual caz în care efectul obţinut este moderat de stilul (aşteptările) asistenţilor de cercetare/terapeuţilor/trainerilor care au implementat studiul; (d)

replicarea studiilor experimentale în contexte variate din perspectiva aşteptărilor cercetătorilor (de exemplu, studiul se replică nu numai de către echipa de cercetători care propun intervenţia respectivă ca o noutate, ci şi de către o echipă concurentă, care poate testa eficienţa intervenţiei respective fără a fi acuzată de conflicte de interese sau de un nivel diferit de aşteptări şi implicare în intervenţiile supuse testării).

3.4.3.3. Validitate de construct la nivelul efectelor (variabilelor) măsurate

În psihologie, cel mai adesea se apelează la constructe abstracte (de exemplu, inteligenţa emoţională) care trebuie operaţionalizate (măsurate) în vederea colectării unor date empirice şi obţinerii unor rezultate. Pe baza rezultatelor obţinute sunt discutate, mai apoi, implicaţiile acestora din nou la un nivel abstract (conceptual, teoretic). De aceea, în linii mari şi indiferent de forma de risc prezentată, atunci când vorbim despre validitatea de construct la nivelul variabilelor măsurate, inclusiv a celor de tipul VD, avem în vedere două întrebări fundamentale: (a) „sunt instrumentul sau procedura utilizate potrivite pentru a descrie constructul?”; (b) „sunt justificabile inferenţele explicative pe care le face cercetătorul de la rezultatele obţinute înapoi la teorie?”.

Prima întrebare se referă explicit la procesul prin care se trece de la constructul teoretic, prin operaţionalizare, la variabila măsurată (indicatorul utilizat) – *de la construct la variabilă, de la general la particular*. Cea de-a doua întrebare se referă la demersul invers, acela de a reuşi să transforme rezultatele (variabilele analizate) în conţinuturi cu sens, relevante teoretic şi conceptual – *de la variabilă la construct, de la particular spre general*.

În secţiunea anterioară, dedicată prezentării riscurilor în legătură cu VI au fost abordate ambele aspecte, cu un accent pe cea de-a doua întrebare, deoarece prin natura lor, studiile experimentale (îndeosebi cele cu un scop cauzal-explicativ) sunt centrate pe oferirea de explicaţii referitoare la mecanismul cauzal implicat. Ambele întrebări sunt importante şi pentru validitatea de construct cu referire la alte variabile decât cele care implică manipulare experimentală (VI). De aceea, o bună parte dintre riscurile discutate în secţiunea precedentă sunt valabile şi pentru variabile-efect (VD) sau pentru variabile provenite din studii non-experimentale. Tabelul 3.9 conţine o descriere a celor mai relevante riscuri privitoare la aceste variabile.

Tabelul 3.9. *Riscuri privitoare la validitatea de construct cu referire la variabilele-efect*

Riscuri privitoare la validitatea de construct a altor variabile decât VI (provenite din studii non-experimentale; VD)
*Complexitatea constructului (*construct confounding*)* *Exemplificare*: Într-un studiu comparativ non-experimental se constată diferenţe între partidele parlamentare şi cele neparlamentare privitoare la probabilitatea de a intra în parlament în mandatul următor, în favoarea celor parlamentare. Asumând că inferenţa este corectă din punctul de vedere al validităţii statistice, pot apărea probleme cu interpretarea (explicarea) rezultatului. De exemplu, cercetătorul ar putea să afirme că partidul parlamentar are mai mari şanse să intre din nou în parlament pentru că pe durata mandatului în care este partid parlamentar are un acces mai mare la resurse, ceea ce poate însemna atragerea unor oameni mai valoroşi/capabili să câştige noi curse electorale. Însă o asemenea inferenţă ar fi problematică din perspectiva validităţii de construct, pentru că partidele parlamentare se deosebesc de cele neparlamentare şi cu privire la alte aspecte cum ar fi: (a) un nivel de mediatizare diferit (timpul de antenă mai ridicat pentru partidele parlamentare, inclusiv în campaniile electorale); (b) o mai bună reţea teritorială (structură organizatorică) pentru partidele cu reprezentanţi în parlament; (c) absenţa impresiei de vot irosit în cazul partidelor parlamentare (de exemplu, în cazul partidelor neparlamentare, deşi unii candidaţi pot fi percepuţi pozitiv, ei nu ajung să fie votaţi pentru că alegătorii nu ar dori să-şi irosească votul respectiv, în condiţiile în care le lipseşte încrederea că partidul respectiv va atinge pragul electoral solicitat).
*Subreprezentarea constructului (*mono-operation bias*)*

Exemplificare: Într-un studiu corelaţional ar fi discutabil să se afirme că există o asociere semnificativă statistic între nivelul de conştiinciozitate şi angajamentul organizaţional din perspectiva validităţii de construct, dacă relaţia se replică doar în cazul în care trăsătura de personalitate vizată este măsurată prin proba X, nefiind susţinută dacă conştiinciozitatea este evaluată prin probele Y sau Z. O problemă similară apare atunci când nivelul de conştiinciozitate corelează cu performanţa în muncă doar în contextul în care criteriul este evaluat prin hetero-raportare, dar nu şi atunci când criteriul este măsurat într-un alt fel (de exemplu, apelând la un indicator obiectiv – criteriu hard). În sfârşit, dacă ar fi să ne raportăm la un exemplu menţionat ceva mai devreme, transferat într-un design non-experimental, privind compararea nivelului de agresivitate la cei care ascultă muzică hip hop *versus* şlagăre, ar creşte validitatea de construct dacă se apelează la un număr variat de artişti şi de piese muzicale din interiorul genului muzical evaluat decât dacă ne-am baza doar pe câte 1-2 piese pentru fiecare din cele două genuri de muzică.

*Confuzia dintre construct şi modalităţile sale (*confounding constructs with levels of constructs*)*
Exemplificare: În cazul în care se constată că nivelul conştiinciozităţii ca dimensiune globală de personalitate corelează semnificativ cu performanţa în muncă, se pot efectua analize ulterioare, la nivel de faţete, pentru a revela care dintre faţetele conştiinciozităţii corelează cu performanţa în muncă (de exemplu, orientarea spre realizare) şi care rămân nesemnificative statistic (de exemplu, prudenţa). Astfel, în cazul studiilor non-experimentale discuţia cu privire la confuzia dintre construct şi modalităţile sale vizează nevoia unei nuanţări în interpretarea/explicarea rezultatelor.

*Distorsiunea metodei comune (*mono-method bias*)*
Exemplificare: Într-un studiu care urmăreşte să identifice potenţialul rol moderator al afectivităţii negative în relaţia dintre conştiinciozitate şi comportamentele contraproductive ale angajaţilor, toate cele trei variabile amintite, importante pentru scopul studiului, sunt măsurate prin aceeaşi metodă – completarea unor probe standardizate ce necesită autoraportare (*self-reports*). Dacă toate variabilele sunt măsurate prin acelaşi tip de metodă (de exemplu, prin autoraportare), metoda devine parte din constructele măsurate, crescând artificial intensitatea asocierii dintre variabile. Aspectele comune pot viza oricare dintre caracteristicile instrumentelor utilizate (probele necesită acelaşi tip de răspuns, pe scală Likert în 5 trepte; probele apelează la autoraportare; probele împărtăşesc acelaşi stil de construcţie a frazei, itemii fiind formulaţi în termeni pozitivi; probele sunt similare din perspectiva nivelului de dezirabilitate socială implicat etc.). Problema apare ca urmare a faptului că nu se poate distinge bine dacă eventuala asociere dintre două variabile este determinată de existenţa unei legături reale între variabile sau este doar un artefact al faptului că ambele variabile împărtăşesc aceeaşi metodă (acelaşi fel de colectare a datelor).

Această problemă este foarte cunoscută în literatura de specialitate din domeniul psihologiei organizaţionale, dar mai puţin cunoscută în domeniul psihologiei clinice, fiind întâlnită sub

denumirea de varianţa metodei comune (*shared method variance* sau *common method bias*) (P.M. Podsakoff, MacKenzie, Lee şi N.P. Podsakoff, 2003). De altfel, într-un studiu ulterior, acelaşi grup de autori semnalau nivelul pandemic al acestui risc (76% dintre studii implică o singură metodă de colectare a datelor, iar 33% dintre studii se bazează exclusiv pe completarea de chestionare autoraportate) (Bodner, 2006 *apud* P.M. Podsakoff, MacKenzie şi N.P. Podsakoff, 2012). Impactul acestei distorsiuni este semnificativ la nivel de rezultate. De exemplu, o metaanaliză a relaţiei dintre factorii de personalitate şi performanţa în muncă, ambele măsurate prin autoraportare evidenţiază un efect mediu r = 0,26, în timp ce aceeaşi relaţie evaluată în contextul utilizării unor metode diferite de colectare a datelor (personalitatea prin autoraportare, performanţa în muncă prin heteroraportare sau prin indicatori obiectivi) conduce la un efect r = 0,11 (Podsakoff *et al.*, 2012).

Dovezi insuficiente sau neconcludente cu privire la validitatea instrumentului utilizat
Exemplificare: Majoritatea aspectelor legate de neajunsuri în calitatea psihometrică a unor instrumente (probleme de consistenţă internă, restricţia de amplitudine etc.) au fost discutate în secţiunea referitoare la validitatea statistică. Însă pot fi întâlnite cazuri în care nu se constată probleme cu privire la acurateţea şi variabilitatea datelor colectate, ci cu privire la semnificaţia acestora sau etichetarea lor. În România, unul dintre exemplele tipice care îmi vin în minte este cazul probei de mobilitate a atenţiei Praga, extrem de întâlnită în laboratoarele şi cabinetele de psihologia transporturilor. Acest test, deşi are o consistenţă internă şi o validitate relativă la criteriu bună, măsoară altceva decât cred cei mai mulţi că măsoară. Astfel, foarte mulţi practicieni cunosc proba ca un instrument de măsurare a atenţiei distributive, confuzie la a cărei origini se află Bontilă (1935). Acesta a etichetat-o ca o probă de atenţie distributivă, fără a oferi sprijin ştiinţific pentru aceasta. La perpetuarea acestei erori a contribuit din plin mediul academic românesc, care a inclus proba în cursurile adresate studenţilor, fără o analiză critică, în categoria testelor de atenţie distributivă. Însă, dacă vom analiza atent conţinutul probei, vom putea identifica imediat faptul că nu există elementele necesare unei sarcini de atenţie distributivă. Aceasta presupune prin definiţie rezolvarea simultană a două sarcini diferite (de exemplu, a număra cu voce tare în sens descrescător, de la 100 la 1, în timp ce se scrie un text nou după dictare).
Situaţii similare pot apărea la case mai mari. De exemplu, în domeniul expansiv al cogniţiilor sociale implicite (Gawronski şi Payne, 2010; Sava, 2011) se vorbeşte despre măsurători implicite ale stimei de sine. Cu toate că există dovezi empirice în sprijinul unei asemenea etichetări, nu există suficiente date pentru a ne asigura că ceea ce se surprinde prin sarcinile respective este stima de sine implicită şi nu concepte apropiate sau similare precum încrederea de sine (*self-confidence*), autoeficacitatea (*self-efficacy*) sau aprecierea de sine (*self-worth*).
Pentru a evita situaţii ce decurg din definirea sau operaţionalizarea inadecvată a unui construct, Trochim (2006) recomandă parcurgerea a trei paşi: (a) definirea constructului prin raportarea la o reţea semantică existentă de constructe (prin ce se aseamănă şi prin ce diferă de alte constructe apropiate ca sens); (b) să operaţionalizeze constructul ţinând cont de aspectele acestuia definitorii

(de exemplu, într-o probă de atenţie distributivă trebuie să se solicite rezolvarea în acelaşi timp a două sarcini diferite, nu doar să se chestioneze individul dacă reuşeşte să facă faţă la mai multe sarcini derulate simultan – care ar fi doar o percepţie autoraportată a nivelului atenţiei distributive, în cel mai bun caz, sau un item specific evaluării nivelului de autoeficacitate, în cazul mai puţin fericit); (c) să existe dovezi empirice care să confirme relaţia dintre constructul evaluat şi alte constructe sau acelaşi construct măsurat prin probe deja consacrate. De obicei, acest punct c lipseşte (sau este mai puţin dezvoltat) în cazul unor probe nou-apărute. De aceea, ar fi indicat ca cercetătorul fie să apeleze la instrumente deja consacrate pentru evaluarea unui anumit construct, fie să apeleze la instrumente mai puţin cunoscute pentru care să se ofere informaţii suplimentare despre validitatea şi fidelitatea probei, pentru a înlătura reticenţele evaluatorilor cu privire la calităţile psihometrice ale instrumentului respectiv.

3.4.4. Validitatea externă

Brewer (2000) subliniază ideea că validitatea externă are în vedere ideea replicării rezultatelor, a generalizării rezultatelor obţinute într-un studiu, pe alte persoane, contexte sau alte proceduri de cercetare (alte modalităţi de manipulare ale VI, alte VD, alte tipuri de design). Din acest punct de vedere, problematica validităţii externe vizează măsura în care rezultatele pot fi replicate în contexte diferite („avem de-a face cu rezultate robuste?”, „ce factori moderatori au fost identificaţi?”) şi cu gradul de realism (transfer) al acestor rezultate în contexte naturale („pot fi efectele transferate în contexte aplicative etc.?”).

3.4.4.1. Replicarea rezultatelor (robusteţea rezultatelor)

Spre exemplu, problema validităţii externe într-un studiu al cărui rezultat principal identifică o relaţie semnificativă statistic între expunerea la

filme cu conţinut violent şi nivelul de agresivitate vizează validitatea inferenţelor cu privire la posibilitatea extinderii acestor rezultate: (a) la alte categorii de participanţi (rezultatele sunt replicate nu numai pe lotul de copii de gimnaziu examinaţi, ci şi pe alte categorii posibile: tineri, adulţi, atât în cazul băieţilor, cât şi al fetelor etc.); (b) în alte contexte decât cel din cercetare (nu doar în studii de laborator, ci şi în contexte naturale, cum ar fi vizionarea unui film violent într-un cinematograf dintr-un mall); (c) la alte modalităţi de operaţionalizare a variabilelor studiate (de exemplu, rezultatul este acelaşi dacă se apelează la un chestionar *self-report* de evaluare a agresivităţii de stare, cât şi dacă se utilizează o măsură indirectă de evaluare a agresivităţii, cum ar fi IAT-ul [testul asocierilor implicite; Greenwald McGhee şi Schwartz, 1998], sau o măsură obiectivă, directă a agresivităţii precum observarea frecvenţei în care este pedepsit un personaj din cadrul unui joc pe calculator ce solicită administrarea de pedepse unui personaj, dacă acesta nu atinge un anumit standard specificat etc.). În sfârşit, dacă studiul este unul experimental, se urmăreşte şi măsura în care rezultatul este replicat în condiţiile în care se variază modul de manipulare a VI (de exemplu, efectul apare în orice tip de material video catalogat drept unul violent, atât în cele care prezintă cazuri ficţionale [filme SF], cât şi în cele care prezintă situaţii reale [„Ştirile de la ora 5”] etc.).

Shadish *et al.* (2002) limitează discuţia despre validitatea externă doar la cazul studiilor experimentale sau cvasiexperimentale, aceasta fiind definită ca măsura în care relaţia cauzală transpare în situaţia varierii persoanelor, contextelor, intervenţiei oferite şi a modalităţii de măsurare a efectelor (VD). Consider această restrângere a sferei de acţiune a discutării aspectelor de validitate externă doar în legătură cu inferenţe referitoare la relaţii de tip cauză-efect ca fiind una limitativă şi mai puţin

inspirată. Aşa cum am amintit anterior, studiile nu vizează numai obiective de tip cauzal-descriptiv sau cauzal-explicativ. De exemplu, ele pot urmări obiective descriptive propriu-zise („care este incidenţa anuală a depresiei majore la adolescenţi?”) sau descriptiv-analitice („poate prezice conştiinciozitatea performanţa la locul de muncă?”). Deşi niciunul dintre aceste studii nu urmăreşte obiective cauzale, ar fi important să vedem în ce măsură rezultatele lor sunt replicate în alte contexte şi/sau generalizabile la o întreagă populaţie de interes. Prin urmare este important să analizăm studiile prin prisma validităţii lor externe. De exemplu, dacă ar fi să ne referim la faptul că a fost obţinută o asociere între conştiinciozitate şi performanţa în muncă, am avea mai mare încredere în şansele de replicare a acestui rezultat într-un context aplicativ, dacă studiul a fost efectuat într-un context natural, ecologic (cum ar fi în cadrul procesului de selecţie), decât dacă datele despre relaţia dintre conştiinciozitate şi performanţă în muncă au fost colectate în cadrul unui studiu de laborator (context artificial), în care performanţa a fost evaluată ca o măsură de stare, prin nivelul atins într-o sarcină de efectuat în cadrul studiului.

De aceea, pentru a analiza nivelul de reproductibilitate a rezultatelor obţinute într-un studiu, un rol important îl prezintă identificarea unor factori moderatori, cu alte cuvinte a efectelor de interacţiune dintre rezultatul obţinut şi diferiţi parametri ai studiului (variaţii cu privire la tipul de participanţi, ale contextului cercetării sau ale variabilelor implicate). Aceste interacţiuni, reprezintă de altfel, principalele riscuri privitoare la validitatea externă, aşa cum reiese din tabelul 3.10, cu precizarea că le vom redefini terminologic faţă de Shadish *et al.* (2002), pentru a nu limita discuţia doar la cazul studiilor ce vizează scopuri cauzale.

Tabelul 3.10. *Riscuri privitoare la validitatea externă (adaptare după Shadish* et al.*, 2002)*

Dependenţa efectului de populaţia studiată Este cunoscut faptul că cele mai multe studii din domeniul psihologiei sunt efectuate pe studenţi, datorită uşurinţei cu care aceştia pot fi selectaţi ca participanţi în studiu. De exemplu, Sears (1986 *apud* Brewer, 2000) prezintă o statistică care indică că nu mai puţin de 74% dintre studiile de psihologie socială publicate în reviste de prestigiu în anul 1985 au fost efectuate pe studenţi din ciclul licenţă. Riscul generat de această stare de fapt cu privire la validitatea externă este de a stabili în ce măsură efectele obţinute în cadrul unei populaţii specifice sunt replicate pe alte categorii de persoane. Validitatea externă priveşte nu numai statutul acestora (sunt replicate efectele pe adulţi, copii, imigranţi, persoane aparţinând unor culturi diferite etc.), ci vizează şi impactul pe care îl are tipul de participanţi selectaţi (voluntari, recompensaţi financiar, preselectaţi etc.). Pentru a înlătura (reduce) asemenea riscuri, se recomandă fie apelul la eşantioane naţionale reprezentative, fie utilizarea unor eşantioane de convenienţă selective (*purposive sampling*), care să includă anumite categorii ţintite de participanţi. Bunăoară, dacă se aşteaptă ca genul persoanei să modereze relaţia dintre variabilele studiate, ar fi bine ca lotul utilizat să includă participanţi de ambele sexe şi să se testeze rolul moderator al genului asupra relaţiei respective.
Dependenţa efectului de contextul cercetării Un efect obţinut poate fi dependent de un anumit context în care se derulează cercetarea. De exemplu, eficienţa pe termen lung a unui program de consiliere în vederea utilizării prezervativelor pentru a reduce riscul de transmitere a bolilor venerice sau al apariţiei sarcinilor nedorite apare doar în studiile implementate în mediul urban, dar nu şi la sate, pentru simplul motiv că prezervativele se găsesc mai uşor la oraş decât la sate. Un alt exemplu ar fi asocierea dintre două variabile, precum observarea relaţiei dintre conştiinciozitate şi comportamentele contraproductive doar într-un context organizaţional în care liderul adoptă un stil de conducere *laissez faire* (pentru că un asemenea şef le oferă oportunităţi celor mai puţin conştiincioşi să devieze mai uşor). Pentru a controla impactul pe care contextul îl are asupra efectului studiat, se poate apela la studii cu un mare număr de participanţi derulate în locaţii multiple (*large multisite studies*). Dintre acestea, studiile randomizate controlate cu un număr mare de participanţi, derulate în locaţii multiple sunt destul de răspândite în domeniul sănătăţii mentale, unde sunt extrem de apreciate. Ele sunt considerate de unii specialişti, precum Turok *et al.* (2011), cel puţin la fel de importante precum metaanalizele de studii experimentale din perspectiva puterii dovezilor empirice obţinute prin asemenea demersuri, în timp ce alţi autori, precum Scifres, Iams, Klebanoff şi Macones (2009) sau LeLorier, Gregoire, Benhaddad, Lapierre şi Derderian (1997), le consideră a fi chiar mai relevante decât rezultatele din metaanalize. Studiile randomizate *multisite* îmbină punctele forte ale unui design randomizat cu consecinţe pozitive asupra inferenţelor legate de validitatea

internă a studiului, cu punctele forte ale unui studiu desfăşurat în locaţii multiple, utile pentru a infera aspecte legate de validitatea externă a concluziilor obţinute. Detalii metodologice şi statistice introductive despre acest gen de studii pot fi găsite în Kraemer (2000) şi Raundenbush şi Liu (2000).

Dependenţa efectului de variaţii ale intervenţiei oferite (de tipul de manipulare a VI)
Efectul VI asupra VD poate fi dependent de anumite caracteristici ale VI. De exemplu, este posibil ca impactul ascultării muzicii hip hop asupra agresivităţii adolescenţilor să apară numai în contextul în care se ascultă anumiţi MC (interpreţi) sau ca efectul coaching-ului executiv asupra performanţei în muncă să apară numai în cazul în care programul se derulează timp de 20 de ore, în locul a 10 ore, în şedinţe de grup, dar nu şi în cele de coaching individual. În aceeaşi categorie a dependenţei efectului de specificul manipulării VI intră interacţiunea dintre VI în cazul unor designuri factoriale. De exemplu, un medicament poate avea un anumit efect dacă este luat singur şi un alt efect dacă este luat în combinaţie cu alte substanţe cu care interferează.

Pentru a analiza impactul situaţiei în care variaţii în modul de operaţionalizare ale VI afectează diferenţiat VD, se poate apela la: (a) analize de moderare în cadrul studiilor metaanalitice, pentru a observa impactul acestor caracteristici asupra intensităţii efectului observat la VD; (b) implementarea unor studii cu un design experimental mai complex (de exemplu, a unui design factorial) pentru a testa direct potenţialul efect moderator; (c) replicarea efectelor obţinute prin varierea caracteristicilor VI şi/sau a modului în care aceasta a fost manipulată (de exemplu, într-un studiu privind inducerea experimentală a unei bune dispoziţii prin tehnica amorsării (*priming*) se poate urmări dacă efectul apare utilizând stimuli-amorsă diferiţi: stimuli vizuali *vs* verbali, cu o durată de expunere de 100 ms *vs* 500 de ms etc.).

Dependenţa efectului de modul de operaţionalizare al variabilelor, altele decât VI
Atunci când se are în vedere studierea unei asocieri între două constructe sau chiar identificarea unei relaţii de tip cauză-efect între acestea, ar fi important să observăm că efectul se replică indiferent de modul în care a fost operaţionalizată variabila măsurată. Însă de multe ori se constată că efectul este unul semnificativ statistic dacă se ia în considerare o anumită variabilă, dar devine unul nesemnificativ atunci când se are în vedere o altă variabilă. De exemplu, într-un studiu privind eficienţa terapiei raţional-emotive comparativ cu tehnicile de relaxare Schultz în reducerea anxietăţii, pot fi găsite diferenţe semnificative statistic dacă se au în vedere scorurile la scala de anxietate Hamilton, dar nu apar diferenţe de eficienţă între cele două forme de psihoterapie dacă se iau în considerare rezultatele la scala de anxietate Zung, cu toate că ambele probe au fost aplicate pe acelaşi lot de subiecţi. Riscul este valabil şi în cazul situaţiilor în care obţinem rezultate diferite pentru aceeaşi asociere studiată, în condiţiile în care s-a apelat la modalităţi diferite de operaţionalizare a variabilelor şi la participanţi diferiţi.
În legătură cu acest tip de risc, merită adus în discuţie exemplul oferit de Heppner, Wampold şi Kivlighan (2008), cu privire la dezbaterea dintre adepţii behaviorismului şi cei ai curentului psihodinamic vizavi de găsirea celor mai eficiente forme de intervenţie. Dacă luăm în considerare

modificările comportamentale (în simptomatologie), răspunsul ar fi că terapiile comportamentale ar fi mai eficiente. Dacă avem în vedere starea de bine (cât de bine se simte pacientul?), situaţia ar putea chiar înclina în direcţia opusă. Astfel, rezultatul este dependent de ceea ce se consideră a fi mai important, fapt pentru care s-a introdus distincţia dintre „a te face mai bine" (*getting better*) şi „a te simţi mai bine" (*feeling better*), diferenţă care are probabil rădăcini filosofice în distincţia dintre abordarea pozitivistă şi abordarea interpretativă.
O metodă utilă de analiză a dependenţei rezultatelor (efectelor) de modul în care au fost operaţionalizate variabilele din studiu este apelul la metaanaliză, un tip de studiu important prezentat pe larg în capitolul 7.

Dependenţa efectului de mediere (mecanismului cauzal) de contextul studiului
Spre deosebire de toate cele patru situaţii anterioare care se focalizau pe identificarea unor factori moderatori, acest risc vizează gradul de generalizare/de replicare a unui factor mediator, ca mecanism cauzal explicativ utilizat pentru relaţia dintre VI şi VD. Riscul privitor la validitatea externă apare atunci când mecanismul cauzal identificat într-un studiu nu se replică într-altul sau se ajunge la identificarea unui mecanism cauzal alternativ. Cu alte cuvinte, efectul obţinut este dependent de mecanisme cauzale diferite, în funcţie de variaţii în contextul (caracteristicile) cercetării.
Un astfel de exemplu ar putea fi cazul condiţionării afective. Prin condiţionare afectivă (*evaluative conditioning* – EC) se înţelege măsura în care un stimul neutru (de exemplu, un cuvânt abstract şi fără sens precum „leinar") primeşte în timp o conotaţie afectivă pozitivă (dacă este asociat repetat cu stimuli necondiţionaţi plăcuţi, cum ar fi fotografii cu bebeluşi râzând) sau o conotaţie afectivă negativă (dacă este asociat repetat cu stimuli necondiţionaţi negativi, cum ar fi persoane suferinde, fotografii cu copii maltrataţi etc.). Aplicaţiile acestui tip de condiţionare sunt numeroase: de la domeniul publicităţii (Sweldens *et al*., 2010) la domeniul clinic (Olantunji, Lohr, Sawchuk şi Westendorf, 2005). O explicaţie (mecanismul cauzal justificativ) oferită pentru acest efect este conştientizarea legăturii repetate dintre stimulul neutru şi cel necondiţionat (De Houwer, 2007). Conform acestui mecanism, valenţa afectivă a stimulului iniţial neutru se va modifica doar după ce participanţii vor realiza (vor conştientiza) legătura repetată dintre stimulul pozitiv şi cel neutru. Alte studii însă contestă acest mecanism, propunând ca mecanism alternativ atribuirea greşită a valenţei afective de la stimulul necondiţionat spre cel neutru, acest lucru fiind realizat în mod inconştient (deoarece efectul a fost obţinut uneori în absenţa conştientizării legăturii dintre stimulul condiţionat şi cel necondiţionat). Prin urmare, ar fi interesant de urmărit în ce context se activează conştientizarea legăturii ca mecanism cauzal explicativ pentru apariţia efectului (pentru modificarea valenţei afective a stimulului condiţionat), respectiv în ce context se activează celălalt mecanism cauzal, atribuirea greşită a afectului. Clarificări aduse în această direcţie de studii ar fi extrem de utile în zona cercetării aplicative, pentru a dezvolta metode eficiente de modificare a atitudinilor în funcţie de mecanismul cauzal activ în situaţia respectivă.

Dependenţa rezultatelor de modalitatea de eşantionare utilizată în selectarea participanţilor la

studiu
Aş adăuga în categoria riscurilor privind robusteţea rezultatelor o problemă clasică întâlnită adesea în studii descriptive (non-experimentale), cum ar fi sondajele de opinie. Selectarea participanţilor prin alte metode de eşantionare decât cele probabilistice (de exemplu, prin apelul la un eşantion de convenienţă), afectează atât validitatea statistică a studiului, cât şi pe cea externă, fiind improbabil să se obţină rezultate reprezentative pentru întreaga populaţie. Cel mai întâlnit exemplu pentru a reflecta acest tip de risc este cel al sondajelor de opinie efectuate de diferite site-uri online sau de canale de televiziune. De exemplu, răspunsurile la întrebarea „Aţi fi de acord cu reluarea demersurilor de suspendare a preşedintelui României, Traian Băsescu?” ar fi diferite dacă am chestiona telespectatorii canalului B1 TV sau dacă am chestiona telespectatorii Antenei 3, iar ambele ar fi distorsionate faţă de răspunsurile populaţiei generale, deoarece niciunul dintre sondajele respective nu apelează la metode probabilistice de selectare a eşantionului de participanţi chestionaţi.
Soluţia tipică pentru acest tip de problemă este selectarea participanţilor la studiu prin metode de eşantionare probabilistice (detalii suplimentare fiind oferite în capitolul 6, în secţiunea destinată descrierii anchetelor sociologice).

3.4.4.2. Reprezentativitatea efectelor în contexte aplicative

Alături de riscurile anterioare ce vizau măsura replicării efectelor, ne putem referi la un tip aparte de dependenţă a efectului de contextul cercetării, când ne referim la validitatea ecologică. Aceasta face referire la măsura în care efectele obţinute în studiu pot fi transferate în lumea reală. Aşa cum se exprima Brewer (2000), rezultatele obţinute pe populaţii atipice (studenţi la psihologie) şi în contexte artificiale (studii de laborator), nu pot fi considerate ca având o validitate ecologică, până când rezultatele se replică în contexte reale, naturale.

Din perspectiva validităţii ecologice sunt mai intens apreciate studiile derulate într-un context cât mai natural (realism mundan – de exemplu, în studierea problematicii jocurilor de noroc prin realitate virtuală sau prin studii derulate într-un cazinou) şi îndeosebi cele care implică aspecte

psihologice similare cu cele care pot fi întâlnite în viaţa reală de către participanţi (realism psihologic)[15].

Din acest punct de vedere, studiile efectuate pot fi analizate din perspectiva gradului lor de similaritate cu realitatea şi/sau cu contextul aplicativ în care se urmăreşte a se analiza dacă s-ar obţine rezultate similare. De exemplu, din perspectiva validităţii externe (ecologice), ar exista o probabilitate mai mare să se replice rezultatul privind eficienţa unei intervenţii educaţionale de prevenire a abandonului şcolar prin implicarea şcolarilor în viaţa comunităţii dacă intervenţia ar fi aplicată pe acelaşi grup-ţintă pe care a fost testată iniţial (copii de imigranţi), decât dacă intervenţia ar viza alte categorii de grup-ţintă, cum ar fi copiii proveniţi din familii defavorizate din mediul rural.

3.4.5. Dinamica relaţiei dintre tipurile de validitate şi tipurile de studii

Aşa cum este o aşteptare nerealistă ca un studiu să-şi propună toată pleiada de obiective posibile (exploratorii, descriptive, cauzale), la fel de nerealist este să ne aşteptăm ca un singur studiu să aibă niveluri ridicate de validitate, în toate cele patru aspecte vizate: statistică, internă, de construct şi externă. De aceea este important să se acorde o prioritate sporită acelor riscuri privitoare la validitatea concluziilor unui studiu în funcţie de obiectivele urmărite.

De exemplu, în cercetările fundamentale, în care accentul se pune pe explicarea fenomenelor psihologice studiate, îndeosebi cu privire la sursa unei relaţii de tip cauză-efect, un rol esenţial îl joacă validitatea de construct, respectiv validitatea internă. În schimb, în cercetările aplicative, în care obiectivul major vizează măsura în care un anumit efect general îşi

găseşte aplicabilitatea în domenii variate, precum în educaţie, în domeniul organizaţional, în sport etc., o importanţă deosebită o are validitatea de construct şi cea externă. De asemenea, prin definiţie, studiile efectuate în condiţii de laborator acordă o mai mare atenţie aspectelor care ţin de validitate internă în dauna celei externe, în timp ce studiile de teren vizează îndeosebi aspectele de validitate externă, existând posibilitatea apariţiei riscurilor (aspecte externe rămase necontrolate) privitoare la validitatea internă.

În sfârşit, cu excepţia validităţii statistice, relevantă pentru orice demers metodologic cantitativ, se poate stabili o oarecare asociere între tipul de obiectiv major vizat de un studiu, anumite particularităţi ale designului său (studii cauzal-descriptive *vs* studii pragmatice) şi tipul de validitate asupra căruia ar trebui să se acorde mai multă atenţie pentru a atinge obiectivul vizat.

Tabelul 3.11. *Relaţia dintre obiectivul vizat de studii şi importanţa diferitelor tipuri de validitate*

	Validitate internă	Validitate de construct	Validitate externă
Obiectiv exploratoriu	*	**	*
Obiectiv descriptiv	*	**	**
Obiectiv descriptiv-analitic (transversal)	*	*	**
Obiectiv descriptiv-analitic (longitudinal)	**	**	**
Obiectiv cauzal-descriptiv (*efficacy*)	***	*	*
Obiectiv cauzal-pragmatic (*effectiveness*)	**	**	***
Obiectiv cauzal-explicativ (mecanisme cauzale)	***	***	*

Legendă: *** rol esenţial; ** rol important; * rol redus.

De exemplu, în studiile cu cel mai înalt nivel de rigurozitate (cele care îşi propun un obiectiv cauzal-explicativ), pentru a oferi o interpretare corectă cu referire la mecanismului cauzal responsabil pentru apariţia efectului, sunt esenţiale atât aspectele ce ţin de validitatea internă („există într-adevăr o relaţie cauzală?”), cât şi cele care privesc validitatea de construct („explicaţia pentru relaţia cauzală este validă?”). În schimb, în studiile care vizează doar identificarea unei relaţii de tip cauză-efect, fără a se centra pe validarea unor mecanisme cauzale responsabile pentru acest fapt, accentul poate fi pus fie pe demonstrarea clară a eficienţei unei intervenţii (prin controlul variabilelor potenţial confundate, în cadrul unor studii cauzal-descriptive – *efficacy*), fie pe demonstrarea clară a faptului că intervenţia este eficientă într-un context real (printr-un control mai redus asupra variabilelor potenţial confundate şi o grijă sporită pentru validitatea ecologică, în cadrul unor studii pragmatice – *effectiveness*).

În schimb, aspectele de validitate internă îşi reduc din importanţă în cazul unor studii care îşi propun obiective de tip descriptiv, deoarece prin definiţie acest obiectiv nu vizează extinderea concluziilor dincolo de aspectul pur descriptiv sau predictiv, în timp ce validitatea internă este centrată spre a identifica care dintre relaţiile descrise sunt de tip cauză-efect. De aceea, în cazul studiilor cu obiective descriptive, ar fi de dorit să se acorde o importanţă mai mare aspectelor care vizează validitatea de construct şi celor referitoare la validitatea externă.

Aceste precizări sunt importante în descrierea limitelor unui studiu şi în importanţa acordată subiectului respectiv în secţiunea de discuţii a unui articol. Astfel, dacă studiul este unul de tip non-experimental, corelaţional, se poate preciza, desigur, că nu se poate stabili clar dacă relaţia identificată este una cauzală sau de alt tip. Ceea ce contează însă este măsura în care relaţia este identificată la nivelul ei real de intensitate

(validitate statistică), faptul că măsurătorile utilizate în studiu surprind corect constructele utilizate (validitate de construct) şi că rezultatele sunt utile, cu valoare aplicativă ridicată pentru a fi implementate în practică (validitate externă, ecologică). În schimb, dacă studiul respectiv este unul experimental, ce urmăreşte un obiectiv cauzal-explicativ, accentul din partea de discuţii este pus pe demontarea eventualelor riscuri privitoare la validitatea internă, respectiv privitoare la validitatea de construct. De exemplu, în cazul în care studiul urmăreşte să testeze dacă cogniţiile iraţionale autodepreciative sunt mecanisme cauzale responsabile pentru eficienţa terapiei raţional-emotive în combaterea depresiei majore este important să arătăm că: (a) nu există alţi factori externi intervenţiei care să se constituie ca explicaţii plauzibile pentru efectul apărut (istoric – alt tratament urmat concomitent, maturare – remiterea spontană a simptomelor etc.); (b) nu există alte explicaţii concurente valide, intrinseci intervenţiei (efectul se menţine dacă controlăm statistic efectul placebo, diferenţele de aşteptări din partea experimentatorului, sau alte tipuri de cogniţii iraţionale măsurate, decât cele autodepreciative).

Ca urmare a celor precizate anterior şi a faptului că în multe situaţii există un cost[16] al creşterii unei anumite forme de validitate, singura soluţie viabilă pentru a răspunde treptat la toate riscurile posibile privitoare la problema investigată este aceea de a iniţia un program de cercetare (un set organizat de studii) menit să răspundă treptat la toate tipurile de risc existente la adresa întrebării de cercetat (Shadish *et al.*, 2002). Cel mai adesea, demersul porneşte de la un studiu care prezintă rezultate interesante (obiectiv descriptiv sau explorator). Acesta este urmat de o serie de studii menite să confirme relaţia şi intensitatea mărimii efectului observat (validitate statistică), iar apoi de studii care să arate că asocierea se bazează pe o relaţie de tip cauză-efect (validitate

internă). Apoi urmează studii care să ofere explicaţii valide pentru relaţia cauzală identificată (validitate de construct) şi care să stabilească cât de robust este efectul şi în ce contexte (limite) apare (validitate externă). Alteori demersul este unul invers, cum ar fi situaţiile în care se observă un efect robust în diferite contexte şi pe diferite populaţii, cu privire la asocierea dintre nivelul de epuizare emoţională şi performanţa profesională (validitate externă). Aceste rezultate pot conduce mai apoi la un program de cercetare în care să se stabilească dacă relaţia este una cauzală (validitate internă), cât de intensă este (validitate statistică) şi cum poate fi ea explicată (validitate de construct). O exemplificare a unui astfel de program de cercetare este oferită în caseta 3.7, în care se combină atât studii de laborator, cât şi experimente efectuate în medii naturale.

În sfârşit, în încheierea secţiunii destinate celor patru tipuri de validitate, trebuie subliniată şi în acest caz logica demersului ştiinţific. Validitatea unui efect (privită în ansamblu) este un proces cumulativ bazat pe studii succesive. Validitatea nu este niciodată demonstrată, ci doar sprijinită prin rezultatele obţinute şi prin măsurile de control implementate prin design.

Caseta 3.7. *O ilustrare a programului de cercetare – relaţia cauzală dintre dispoziţia afectivă şi altruism (stările pozitive inducând o creştere a altruismului; adaptare după Brewer, 2000)*

Programul debutează cu o constatare rezultată dintr-un studiu empiric corelaţional: indivizii cu o dispoziţie afectivă mai bună şi cei mai optimişti obţin scoruri mai ridicate la o scală de altruism, aşadar între cele două există o legătură. O parte importantă din program este destinată derulării unor studii experimentale în condiţii de laborator cu scopul de a identifica cu claritate faptul că o dispoziţie afectivă pozitivă induce o creştere a nivelului de altruism. Pentru a da robusteţe acestor rezultate, studiile experimentale variază modalitatea de manipulare a VI: în unele studii participanţii citesc pasaje de text, unele vesele, altele triste; în alte situaţii ascultă muzică pop energizantă cu mesaj

optimist sau sunt alocaţi audierii unor melodii melancolice, lente, cu mesaj trist; iar în alte situaţii primesc un feedback pozitiv sau unul negativ. O situaţie similară apare şi cu referire la VD măsurate, acestea variind de la scale consacrate de măsurare a nivelului de altruism la situaţii experimentale induse – generozitatea financiară dovedită faţă de o cauză aparent nerelaţionată cu studiul (ajutarea sinistraţilor în urma unor inundaţii) sau disponibilitatea de a oferi ajutor unui complice al experimentatorului.

Toate aceste variaţii înregistrate în studiile experimentale de laborator converg, susţinând că dispoziţia afectivă pare a modifica nivelul de altruism. Cu toate acestea, unele critici aduse experimentelor de laborator, scot în evidenţă riscuri precum: (a) caracterul artificial al unor sarcini; (b) situaţia caracteristicilor sarcinii, aspecte ce pot afecta validitatea externă a rezultatelor, lipsind dovezi cu privire la validitatea ecologică.

Pentru a elimina aceste critici în contextul păstrării ideii manipulării VI, esenţială pentru studiile care vizează obiective cauzale, a fost dezvoltat un cvasiexperiment derulat în context natural. Astfel, o parte dintre participanţi (studenţi) a vizionat un film evaluat într-o etapă de pretestare pe un alt lot de studenţi ca fiind trist, iar o altă parte dintre participanţi (studenţi) a vizionat un film care în pretestare a fost evaluat ca fiind neutru (lotul de control). La ieşirea de la cinematograf erau amplasate urne ale Crucii Roşii, unde participanţii puteau dona bani pentru o cauză nobilă. Suma era la latitudinea fiecăruia, nivelul de generozitate financiară constituind VD a studiului. O problemă cu acest studiu a fost distorsiunea de selecţie (participanţii şi-au ales singuri ce film vor să vizioneze, ei nefiind distribuiţi randomizat în cele două condiţii, pentru a reduce la minimum eventualul impact al reactivităţii participanţilor). Astfel, până după momentul părăsirii locaţiei unde se putea dona pentru Crucea Roşie, situată la ieşirea de la film, aceştia nu bănuiau că sunt participanţi într-un studiu psihologic. Mai mult, pentru a controla eventuale diferenţe în privinţa nivelului de altruism existent de la început între cele două grupe constituite după principii neprobabilistice, în anumite seri locaţia unde se putea dona pentru Crucea Roşie era amplasată la intrarea la film. Datele au arătat că, deşi nu erau diferenţe semnificative statistic între cele două loturi înainte de vizionarea filmelor, ele există şi replică rezultatele obţinute în laborator dacă analiza se face pentru situaţiile în care donarea avea loc la ieşirea de la film.

Toate aceste studii converg spre a susţine relaţia cauzală dintre dispoziţia afectivă şi nivelul de altruism, moment în care programul de cercetare poate trece la următorul nivel, de oferire clară a unor dovezi că dispoziţia afectivă şi nu alte constructe asociate acesteia este într-adevăr responsabilă (explicaţia) pentru modificarea nivelului de altruism.

3.4.6. Problema validităţii în studiile bazate pe o metodologie calitativă

Toate aspectele tehnice discutate în secţiunile precedente, de la 3.4.1 la 3.4.5, vizează diferite tipuri de validitate aplicabile în cadrul studiilor empirice bazate pe o metodologie cantitativă (cu accent pe studiile experimentale). În cazul studiilor care utilizează o metodologie calitativă, problema validităţii inferenţelor formulate pe baza datelor colectate rămâne de actualitate, chiar dacă ea este conceptualizată altfel. În acest context, discuţia nu se pune în termenii clasici ai celor patru tipuri de validitate, ci prin strategiile specifice metodelor calitative de validare a inferenţelor formulate, cu toate că şi în acest caz putem vorbi despre interpretări alternative, replicarea rezultatelor, distorsiuni şi artefacte.

Primul lucru care trebuie înţeles de toată lumea este că metodologia calitativă are o logică diferită de cea cantitativă, de aceea nu trebuie judecată după criterii de evaluare specifice validităţii din studiile cantitative. Kazdin (2010) afirma că, spre deosebire de metodologia cantitativă care este interesată să simplifice lucrurile (câţi participanţi sunt, ce răspunsuri medii au obţinut, dacă sunt diferenţe între grupele analizate în medii controlate), metodologia calitativă urmăreşte mai degrabă să dezvolte (să complice) lucrurile, pentru a înţelege fenomenul studiat aşa cum apare el în contextul său natural. Termeni precum semnificaţie, înţelesuri, reprezentări, care împărtăşesc toate ideea unor evaluări subiective, constituie puncte centrale într-un demers calitativ. Această modalitate de a face cercetare poate părea că induce ambiguitate şi un subiectivism excesiv, unei persoane obişnuite exclusiv cu abordarea cantitativă.

Pe de altă parte, deşi există un grup minoritar de autori care fac o opinie separată, cei mai mulţi, la care subscriu şi eu, consideră că studiile calitative pot fi foarte utile numai în anumite scopuri de cercetare, cu precădere să ajute la descrierea (scop descriptiv) sau la înţelegerea şi interpretarea fenomenelor studiate (scop exploratoriu). În privinţa obiectivelor cauzale, studiile calitative pot fi utile între anumite limite: (a) fie pentru a genera teorii (explicative) şi ipoteze de cercetare care să fie verificate prin studii cantitative; (b) fie ca metode complementare celor cantitative, în cadrul unor metodologii mixte menite să ajute la găsirea de explicaţii pentru relaţiile cauzale analizate (ca urmare a descrierilor bogate obţinute de la participanţi şi a interpretărilor şi semnificaţiilor oferite de către aceştia).

Ca urmare a acestor precizări rezultă şi cele cinci tipuri de validitate specifice unei metodologii calitative (Maxwell, 1992), sintetizate în tabelul 3.12. Dintre acestea, primele trei reprezintă tipuri primare de validitate (importante pentru validitatea inferenţelor din studiu), iar ultimele două reprezintă tipuri secundare de validitate (importante mai degrabă pentru a creşte încrederea sau percepţia că rezultatele studiului sunt valide). În categoria tipurilor secundare de validitate ar mai putea fi incluse aspecte calitative precum bogăţia, expresivitatea sau profunzimea descrierilor şi interpretărilor (Whittemore, Chase şi Mandle, 2001).

Tabelul 3.12. *Scurtă explicitare a tipurilor de validitate întâlnite în studiile calitative*

Validitate descriptivă Vizează măsura în care au fost colectate (înregistrate) corect datele oferite de participanţi. Este nivelul cel mai simplu de validitate implicat şi singurul care priveşte realitatea factuală, nu pe cea construită. Are în vedere descrierea acurată a faptelor povestite, aşa cum s-au întâmplat acestea. *Exemplu ipotetic*: participantul a declarat că a aruncat cu un pet de plastic după profesor. *Mijloace de validare*: de verificat pe înregistrările audio dacă informaţia transcrisă este una corectă (nu au fost distorsionate sau omise elemente descriptive importante). *Riscuri*: codarea eronată a datelor; probleme în colectarea datelor.

Validitate interpretativă Vizează acurateţea semnificaţiei acordate actului descris. Sunt oferite interpretări corecte pornind de la opiniile, intenţiile şi trăirile descrise de participanţi. Vizează aspecte ale realităţii construite pornind de la date factuale. *Exemplu ipotetic*: participantul a recurs la acest gest ca un semn de frondă (rezistenţă) faţă de încercările profesorului de constrângere a acestuia. *Mijloace de validare*: interpretarea este validă dacă respectă perspectiva oferită de actorii implicaţi şi dacă, preferabil, se atinge consensul (participanţii oferă aceeaşi interpretare gestului respectiv). *Riscuri*: similare cu cele întâlnite în operaţionalizarea constructelor teoretice.
Validitate teoretică Se oferă explicaţii abstracte pentru comportamentul descris şi interpretat prin apelul la diverse teorii sociale sau psihologice, mergându-se dincolo de perspectiva participanţilor. Este nivelul care implică cel mai mult capacitatea cercetătorului de a extrage esenţialul, de a sesiza patternuri care să conducă la insighturi valoroase despre fenomenul studiat, de a face legături între datele colectate şi anumite modele teoretice care facilitează descrierea, înţelegerea, interpretarea şi explicarea fenomenului studiat. *Exemplu ipotetic*: aruncarea petului înspre profesor semnifică un act de frondă din partea elevului, acest comportament făcând parte dintr-o simptomatologie mai largă ce se referă la degradarea continuă a statutului profesorului în societatea actuală românească şi la o criză a valorilor învăţate în şcoala de astăzi. *Observaţie*! Prima parte din exemplu, aruncarea petului ca act de rezistenţă, poate fi perceput ca un demers invers întâlnit în validitatea de construct (rezistenţa fiind constructul vizat, iar actul de aruncare a petului, modalitatea de operaţionalizare). În schimb, cea de-a doua parte a exemplului (rezistenţa ca parte a unei crize majore a sistemului educaţional actual, a degradării statutului profesorului) ar putea fi analizată din perspectiva validităţii interne, astfel că preocuparea de bază devine aceea de a elimina alte posibile „cauze” care să explice rezistenţa elevului. Acest lucru pare la prima vedere un demers eminamente speculativ, însă el poate fi corectat de către cercetător prin punerea în comun (triangulare) a celorlalte informaţii de care dispune, câteva zeci de pagini de transcrieri a interviurilor de profunzime efectuate. *Mijloace de validare*: explicaţia este validă dacă (a) este recunoscută de către participanţi ca o cauză posibilă, preferabil atingându-se consensul, (b) teoria utilizată pentru a explica fenomenul studiat este fundamentată empiric (există dovezi în sprijinul acesteia). *Riscuri*: distorsiunea indusă de cercetător, omiterea datelor discrepante faţă de modelul teoretic propus, omiterea unor explicaţii alternative.
Validitate externă (generalizare) Prin aceasta se urmăreşte măsura în care modul de descriere, interpretare şi/sau explicare a fenomenului studiat poate fi relevant (valabil) şi pentru alte persoane sau contexte. De reţinut însă că nu sunt căutate interpretări valabile pentru toţi participanţii la studiu, ci pentru cât mai mulţi

dintre ei, recunoscându-se unicitatea şi viziunea individuală a fiecăruia. În acest sens, generalizarea poate avea în vedere măsura în care explicaţia teoretică utilizată are sens, nu numai pentru un alt participant, ci şi pentru alte comportamente specifice (decât aruncarea petului). Ideea generalizării în studiile calitative nu se rezumă doar la aspecte tehnice precum obţinerea unor rezultate similare pe alţi participanţi sau utilizând alte metode de cercetare pentru colectarea datelor, ci urmăreşte, îndeosebi gradul de extindere (valabilitate) a interpretărilor şi a explicaţiilor teoretice oferite. *Exemplul ipotetic*: aruncarea petului este numai unul dintre actele de frondă făcute; la acesta se adaugă distrugerea mobilierului şcolar, negocierea notelor cu profesorul, preluarea apelurilor telefonice în timpul orei; toate acestea converg către ideea unei crize majore a sistemului educaţional de azi. *Mijloace de validare*: măsura în care în urma studiului, aspectele generale (abstracte) sunt potrivite pentru a descrie contexte specifice, în care situaţii particulare pot fi descrise, interpretate sau explicate prin apelul la un nivel categorial superior (aruncarea petului – categorie inferioară *vs* act de indisciplină – categorie superioară) sau mai abstract de analiză şi care să fie relevant pentru cât mai mulţi participanţi. *Riscuri:* reproductibilitate scăzută a rezultatelor şi interpretărilor formulate.
Validitate evaluativă Prin aceasta se urmăreşte validitatea interpretărilor pe care le face cercetătorul faţă de comportamentele observate. Astfel, dacă validitatea interpretativă oferă perspectiva participanţilor cu privire la semnificaţia comportamentelor descrise, validitatea evaluativă se referă la punctul de vedere al cercetătorului, la semnificaţiile pe care el sau ea le acordă comportamentelor observate. Acest tip de implicare subiectivă a cercetătorului, inclusiv la nivelul acordării unor semnificaţii diferite de cele ale participanţilor, este o acţiune legitimă în cercetarea calitativă. *Exemplu ipotetic*: gestul elevului este un strigăt de ajutor, o modalitate de semnalizare a prezenţei sale şi a faptului că a fost neglijat de către profesor. *Mijloace de validare*: nivelul de suport existent în datele colectate pentru aprecierile subiective făcute. *Riscuri*: similare cu cele specifice validităţii de construct; transfer şi proiecţii personale, fără legătură cu datele colectate.

Alături de aspectele menţionate anterior, în aprecierea studiilor calitative au fost introduse o serie de concepte relevante pentru problematica validităţii. De exemplu, conceptul de triangulare se referă la utilizarea unor proceduri sau surse multiple de informare care converg pentru a susţine aceeaşi concluzie. Utilizarea ei ca tehnică de lucru creşte

probabilitatea validării inferenţelor pentru toate tipurile de validitate, oferind mai multă *plauzibilitate* descrierilor, interpretărilor şi explicaţiilor oferite. Strâns legat de acest concept este cel de *credibilitate*. O interpretare este credibilă dacă este făcută într-o manieră coerentă, cursivă şi în care, atât participanţii, cât şi cercetătorul, se regăsesc ca descriere. În sfârşit, un alt concept important este cel de reproductibilitate, care vizează măsura în care un alt cercetător decât cel implicat ar putea ajunge la aceleaşi concluzii, ca urmare a unui demers transparent şi sistematizat de colectare şi analiză a datelor (Kazdin, 2010).

Împreună, indiferent de nivelul de validitate vizat, plauzibilitatea, credibilitatea şi reproductibilitatea, contribuie semnificativ la considerarea inferenţelor formulate ca fiind unele valide[17].

3.5. O perspectivă pragmatică asupra ciclului natural de cercetare

În linii mari, în marketing există o linie directoare clară care porneşte de la analiza nevoilor clienţilor şi identificarea unei nevoi de satisfăcut, continuată cu dezvoltarea unui produs/serviciu care să răspundă acestora, cu stabilirea unei strategii de marketing legate de lansarea acestuia şi apoi de strategii de câştigare a unei cote cât mai mari de piaţă.

Acelaşi lucru se întâmplă şi în domeniul tehnologiei. Ca urmare a extinderii cunoaşterii ca rezultat al cercetărilor fundamentale, poate fi concepută şi mai apoi testată eficienţa unei inovaţii menite să amelioreze sau să optimizeze o anumită tehnologie, prin care să se răspundă mai bine nevoilor existente pe piaţă. După numeroase testări sunt eliminate problemele identificate şi este realizat prototipul final. Acesta este prezentat clienţilor interesaţi direct, care pot decide să treacă la producţia de serie.

În psihologie nu există obişnuinţa de a se gândi în termenii unei priviri de ansamblu, cum ar fi aceea de a urmări etapele cunoaşterii de la identificarea unei probleme psihologice până la stadiul final de diseminare şi oferire de servicii, incluzând cele mai eficiente soluţii identificate pentru problema respectivă. Între motivele care contribuie la această situaţie este faptul că un ciclu complet (de la cercetare fundamentală, la cercetare aplicativă, iar mai apoi la cercetare pentru dezvoltare – inovare) durează semnificativ mai mult (uneori o viaţă de om) decât situaţia tipică legată de ciclul de viaţă al unui produs sau serviciu în marketing sau în inginerie. Un alt motiv îl constituie relativa

focalizare a cercetătorilor, fie în zona cercetării fundamentale, fie în cea a cercetării aplicative.

Însă absenţa unei gândiri pragmatice cu privire la ciclul natural al cercetării pe o anumită tematică din psihologie nu ar trebui să persiste, datorită beneficiilor multiple pe care le aduce o astfel de perspectivă pragmatică de ansamblu. Din acest punct de vedere, figura 3.7 sintetizează toate caracteristicile principale urmărite printr-un astfel de demers, având drept fir logic de start focalizarea pe cunoaşterea problemei investigate (cel mai adesea adresabilă prin cercetarea fundamentală), iar ca punct final demersuri specifice cercetării aplicative şi cercetării pentru dezvoltare-inovare (focalizarea pe identificarea de soluţii care să îmbunătăţească starea de fapt).

Din punctul de vedere al tipului de studii efectuate, în partea stângă a figurii 3.7 se regăsesc într-o mai mare măsură studii care urmăresc obiective exploratorii sau descriptive, deoarece procesul de cunoaştere evoluează gradual dinspre clarificarea problemei investigate înspre identificarea mecanismelor cauzale responsabile pentru apariţia ei. Obiectivele cauzal-descriptive, respectiv cauzal-explicative se pot întâlni în întregul ciclu de studiu al unei problematici, fiind însă măsuri obligatorii pentru studiile focalizate pe intervenţii (soluţiile oferite – partea din dreapta a figurii 3.7). Aşadar, logica ştiinţifică apelează adesea la un asemenea demers, întâi apărând preocuparea pentru a conceptualiza fenomenul studiat, iar mai apoi de a găsi soluţiile aplicative potrivite care să amelioreze problema constatată.

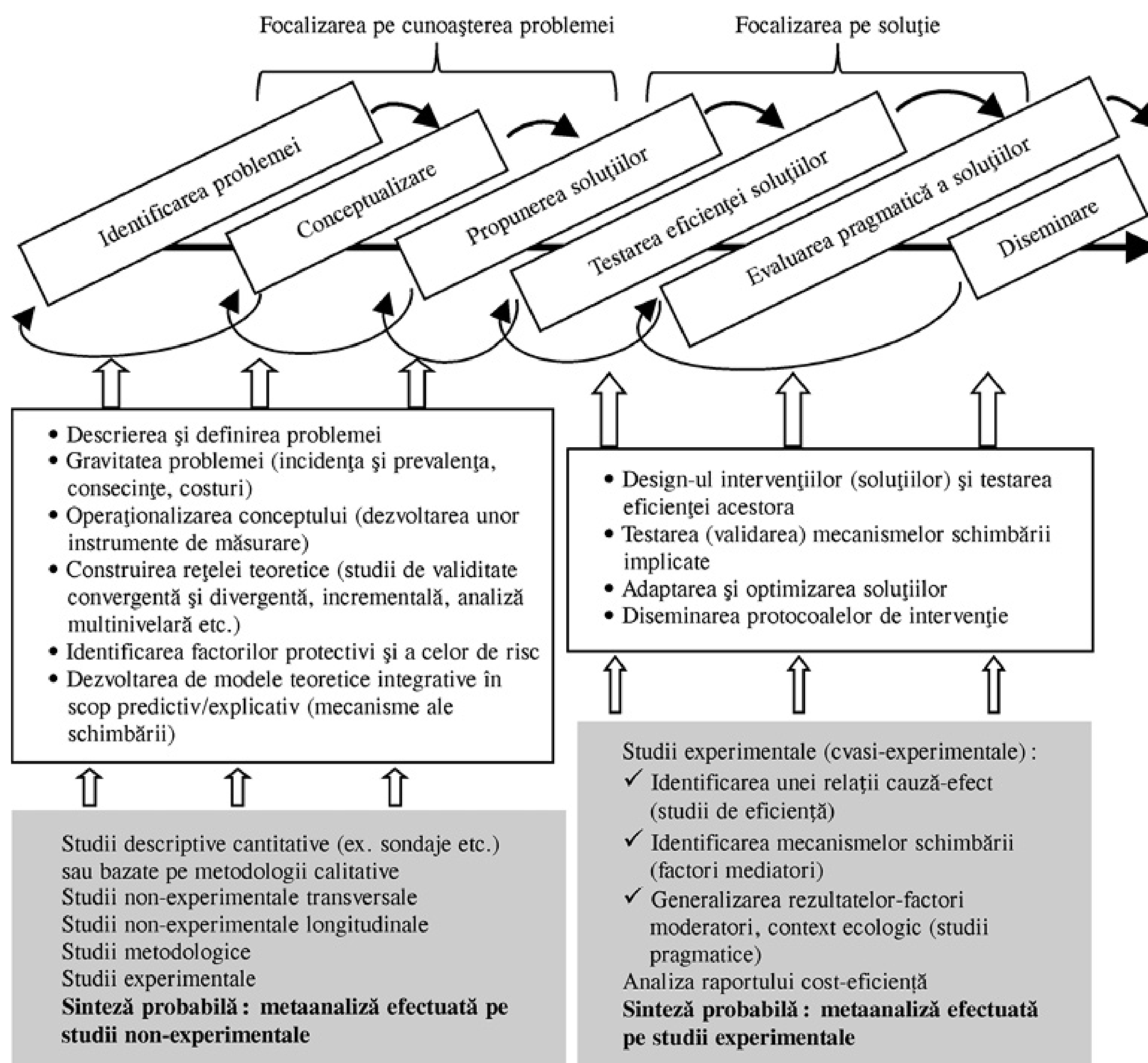


Figura 3.7. *Ciclul ştiinţific natural al studierii unui aspect psihologic – o viziune pragmatică*

Pentru a înţelege ciclul de cunoaştere al unei tematici, voi ilustra prin tema epuizării emoţionale (*burnout*) descrisă pentru prima dată în anii ’70, care a cunoscut o expansiune extraordinară (peste 6.000 de publicaţii; Schaufeli, Leiter şi Maslach, 2009).

Tabelul 3.12. *Ilustrarea aspectelor centrale din ciclul natural ştiinţific în cazul problematicii epuizării emoţionale (*burnout*)*

Aspecte generale (obiective vizate)	**Ilustrare la nivelul studierii epuizării emoţionale**
Descrierea şi definirea problemei	Conceptul de *burnout* a fost introdus la mijlocul anilor

(*obiective exploratorii sau obiective descriptive propriu-zise*) Cele mai întâlnite tipuri de studii care răspund acestor nevoi sunt studiile de caz şi studiile bazate pe o metodologie calitativă. Aceste demersuri sunt menite să faciliteze înţelegerea fenomenului studiat şi să identifice atributele caracteristice, necesare şi definitorii pentru fenomenul semnalat. Această etapă este obligatorie pentru orice construct studiat, fiind faza în care se ajunge (de cele mai multe ori) la un consens cu privire la definirea şi sfera de includere a constructului respectiv.	’70 de Maslach (1976), respectiv Freudenberger (1975), ca o primă încercare de sistematizare a unei serii de studii de caz şi a unor observaţii din practica organizaţională, conform căreia o parte din angajaţi ajung să fie extenuaţi mental şi să le scadă nivelul de implicare în muncă, aceasta pierzându-şi din semnificaţie. Pe baza acestor studii exploratorii şi descriptive s-a ajuns la o definire a conceptului (Maslach, 1981) şi la o descriere a procesualităţii fazelor de burnout (Freudenberger, 1982 *apud* Kaschka, Korczak şi Broich, 2011).
Extinderea şi gravitatea problemei (*obiective descriptive propriu-zise sau descriptiv-analitice*) Într-o societate care acţionează în baza unui model cu resurse limitate ar fi ideal să se stabilească gradul de răspândire a problemei identificate şi direcţia evoluţiei acesteia (de exemplu, o problemă tot mai des întâlnită) prin studierea prevalenţei fenomenului (numărul de cazuri existente) şi a incidenţei acestuia (numărul de cazuri nou-apărute), precum şi a consecinţelor (costurilor) la nivel social, economic etc. Cele mai întâlnite tipuri de studii din această sfera sunt studiile non-experimentale descriptive, cum ar fi sondajele de opinie, studiile corelaţionale sau analiza de costuri.	Problema este destul de răspândită în populaţia generală de angajaţi, fiind observată o prevalenţă a fenomenului la un grad de severitate ridicat (*burnout* clinic) cuprins între 4% şi 7%, ce poate ajunge chiar la 10% în profesiile care presupun un contact uman susţinut (cadre didactice, medici de familie, comercianţi etc.) (Schaufeli, 2003). Prin gravitatea consecinţelor existente la nivel personal, *burnoutul* a fost inclus în ICD-10, iar prin consecinţele negative la nivel de organizaţie (fluctuaţie de personal, absenteism, scăderea satisfacţiei în muncă) a devenit o preocupare curentă pentru specialiştii în resurse umane (Kaschka *et al.*, 2011).
Operaţionalizarea conceptului *(obiective descriptiv-analitice)* În această categorie intră studiile metodologice centrate pe dezvoltarea,	Instrumentul de bază utilizat pentru a măsura nivelul epuizării emoţionale este MBI (*Maslach Burnout Inventory*) (Maslach şi Jackson, 1986 *apud* Schaufeli, 2003). Acesta are trei versiuni: una pentru cadre

adaptarea sau rafinarea unor instrumente psihologice necesare pentru a măsura constructul evaluat (studii descriptiv-analitice). Ele se află în strânsă legătură cu definirea şi descrierea problemei, respectiv cu stabilirea şi identificarea unei reţele (conexiuni) teoretice între constructul examinat şi alte constructe relevante.	didactice, alta pentru alte profesii ce implică contact uman frecvent (oferirea de servicii), iar cea de-a treia versiune, generală, pentru alte tipuri de slujbe. Toate acestea redau cele trei dimensiuni majore ale conceptului: cel de extenuare, cel de depersonalizare şi cel de realizări personale scăzute (ineficienţă). Deşi au fost dezvoltate alte instrumente concurente, multe dintre acestea sunt focalizate doar pe surprinderea extenuării ca o subcomponentă a epuizării emoţionale, neglijând partea care vizează distanţarea faţă de muncă (prin depersonalizare şi ineficienţă).
Construirea reţelei teoretice (*obiective descriptiv-analitice*) În această categorie intră studiile orientate spre construirea reţelei nomotetice dintre constructul analizat şi alte constructe relevante. De exemplu, studiile orientate spre validarea incrementală, cele de validitate convergentă sau divergentă, intră în această categorie. Tot aici s-ar include şi studiile non-experimentale focalizate pe identificarea de moderatori sau mediatori ai relaţiei dintre constructul măsurat şi alte variabile.	Maslach, Schaufeli şi Leiter (2001) trec în revistă o serie de distincţii importante între *burnout* şi alte constructe similare cum ar fi stresul la locul de muncă, suprasolicitarea, insatisfacţia legată de locul de muncă sau depresia. Opusul ideii de *burnout* ar fi cea de angajament organizaţional. În sfârşit, se stabilesc relaţii între *burnout* şi o multitudine de variabile din zona organizaţională precum ambiguitatea de rol, conflictul de rol, relaţiile cu colegii şi cu superiorii, nivelul de absenteism, satisfacţia la locul de muncă, comportamentele contraproductive etc. Aceeaşi autori subliniază că stima de sine scăzută sau un locus al controlului extern joacă in rol moderator, facilitând apariţia *burnoutului*.
Identificarea factorilor protectivi şi ai celor de risc (*obiective descriptiv-analitice şi obiective cauzal-descriptive*) Această direcţie de studii poate fi inclusă în categoria construirii unor reţele teoretice, însă am preferat abordarea ei distinctă pentru a sublinia faptul că se are în vedere găsirea acelor determinanţi (factori) care pot facilita sau inhiba apariţia fenomenului respectiv. Printr-un asemenea demers predictiv, situat la graniţa dintre un demers descriptiv şi	În acest caz se apelează îndeosebi la studii non-experimentale longitudinale, care permit identificarea clară a ordinii în succesiunea trăirilor psihice. De exemplu, este *burnoutul* o consecinţă a unei satisfacţii scăzute la locul de muncă sau este o cauză care contribuie la scăderea satisfacţiei legată de muncă? *Burnoutul* are drept consecinţe o serie de manifestări negative precum depresia, anxietatea etc. sau este un rezultat al acestor stări? În pofida importanţei acestor răspunsuri, există puţine studii longitudinale pe această temă (Schaufeli, 2003). Ele par a sugera o cauzalitate reciprocă în cazul aspectelor de sănătate

unul cauzal, pot fi identificate categoriile vulnerabile (persoane care prezintă multipli factori de risc şi nu dispun ca resurse de factorii protectori) şi se pot testa anumite presupoziţii cu privire la cauzele apariţiei fenomenului respectiv, cu limitările metodologice de rigoare. Studiile non-experimentale transversale, dar mai ales cele longitudinale, fie ele de tip corelaţional sau comparativ, sunt studiile utilizate cel mai des pentru a surprinde potenţialii moderatori sau mediatori ai fenomenului studiat.	(stările negative facilitează *burnotul*, care la rândul său întreţine o stare psihică negativă) şi induc o serie de consecinţe negative la nivel organizaţional (insatisfacţie profesională). Studiile de tip longitudinal pot induce însă în eroare din cauza naturii cronice a *burnoutului*. În lipsa unei variabilităţi în scorurile de la MBI, îndeosebi în cazul operaţionalizării de natură categorială (prezent *vs* absent), pot apărea eronat, asocieri nesemnificative statistic cu alţi factori de risc sau protectori. Alternativa la studiile longitudinale vizează implementarea unor modele structurale (de exemplu, analiza de cale) în cadrul unor studii non-experimentale transversale, prin care o serie de caracteristici ale locului de muncă (de exemplu, autonomia decizională etc.) au fost conceptualizate ca antecedenţi ai *burnoutului*, în timp ce alţi factori (de exemplu, atitudinea negativă faţă de locul de muncă) sunt percepute drept consecinţe ale acestui fenomen.
Dezvoltarea de modele teoretice integrative în scop predictiv/explicativ (analiza mecanismelor schimbării) *(orice tip de obiectiv, preferabil obiective de tip cauzal-explicativ)* Pe măsura acumulării de cunoştinţe despre fenomenul studiat, îndeosebi a dezvoltării reţelei de constructe teoretice şi a determinanţilor (factori de risc, factori protectori) pentru efectul observat, se pot dezvolta modele sau teorii care să prezică şi/sau să explice evoluţia fenomenului studiat. Cel mai adesea, pentru a testa asemenea modele se apelează la studii experimentale ce-şi propun scopuri cauzal-explicative, însă atunci când acest lucru nu este posibil sunt luate în considerare dovezi indirecte provenite din alte tipuri de studii.	Din punct de vedere teoretic au fost dezvoltate mai multe modele explicative pentru apariţia *burnoutului* (la nivel individual, la nivel interpersonal, la nivel organizaţional şi la nivel societal), însă pentru foarte puţine dintre acestea există un suport empiric direct (Schaufeli, 2003). Una dintre cele mai cunoscute conceptualizări vede epuizarea emoţională ca o consecinţă a inegalităţii schimburilor de la locul de muncă (angajaţii resimt că au dăruit/s-au implicat prea mult, comparativ cu cât au primit înapoi, fiind vorba astfel de un efort nerăsplătit, indiferent dacă avem în vedere şeful direct sau întreaga organizaţie ca sursă a insatisfacţiei) (Schaufeli, 2003). Un alt model celebru a fost dezvoltat de Demerouti *et al.* (2001), *burnoutul* fiind considerat rezultatul unui dezechilibru între solicitările de la locul de muncă şi resursele psihologice disponibile pentru a face faţă la acestea. Presiunea indusă de solicitări conduce îndeosebi la extenuare, în timp ce lipsa resurselor contribuie

	îndeosebi la dezangajare în sarcină (scăderea implicării).
Designul intervenţiilor şi testarea eficienţei acestora *(obiective cauzal-descriptive*) În urma acumulării unui număr mare de cunoştinţe despre problema studiată, pot fi concepute şi evaluate o serie de programe menite să diminueze sau să elimine problema respectivă. Modalitatea standard de testare a eficienţei intervenţiilor propuse este apelul la studii cauzal-descriptive (*efficacy studies*).	Van der Klink, Blonk, Schene şi Van Dijk (2001) au efectuat o metaanaliză privind eficienţa intervenţiilor menite să reducă nivelul de *burnout*, rezultate încurajatoare fiind obţinute pentru programele cognitiv-comportamentale şi, secundar, pentru intervenţiile bazate pe tehnici de relaxare. Totuşi, se observă că cele mai multe intervenţii utilizate nu sunt specifice exclusiv *burnoutului*, ci vizează aspecte mai generale cum ar fi combaterea depresiei, a stresului etc. (Schaufeli, 2003).
Identificarea mecanismelor schimbării *(obiective cauzal-explicative)* Dacă împreună cu testarea eficienţei unei intervenţii se iau măsuri suplimentare de a identifica ingredientul activ, responsabil pentru apariţia schimbării, se urmăresc obiective cauzal-explicative. Modalitatea standard de testare a acesteia presupune studii cauzal-descriptive, având un design special care să permită identificarea mecanismelor schimbării.	Nu au fost identificate studii care să includă explicit măsuri de testare a unor mecanisme cauzale odată cu testarea eficienţei unei intervenţii la nivel de *burnout*. De fapt, de cele mai multe ori *burnoutul* a fost utilizat ca un construct de tip „umbrelă” pentru a explica relaţia dintre un factor antecedent (de exemplu, ambiguitatea de rol) şi o consecinţă (de exemplu, insatisfacţia profesională).
Adaptarea şi optimizarea soluţiilor *(obiective cauzal-descriptive)* În urma derulării studiilor cauzal-descriptive şi a celor cauzal-explicative, identificăm cu claritate dacă intervenţia este una eficientă şi care sunt ingredientele active (mecanismele responsabile pentru eficienţă). Însă aceste rezultate nu asigură faptul că intervenţia este eficientă în orice context şi pe participanţi neselectaţi special. Nu sunt clare nici limitele eficienţei intervenţiei (variabilele moderatoare). De aceea, pentru a răspunde acestor	Majoritatea studiilor efectuate pentru a testa eficienţa intervenţiilor menite să reducă nivelul de epuizare emoţională se constituie, în realitate, ca studii pragmatice, derulate în mediul natural al participanţilor şi fără criterii dure de includere/excludere a celor chestionaţi din analiză. Aşadar se poate spune că intervenţiile respective sunt eficiente inclusiv în condiţii asemănătoare cu cele din realitate. Dintre intervenţiile analizate, cele care combină intervenţii la nivelul individului cu măsuri ameliorative la nivel organizaţional conduc la efecte benefice pe o perioadă mai lungă de timp decât cele

neajunsuri se apelează la studii pragmatice (*effectiveness studies*). Prin acest gen de studii se are în vedere evidenţierea robusteţii eficienţei intervenţiei în contexte naturale variate, pornind de la premisa că este importantă reproducerea efectelor în condiţii asemănătoare cu cele din realitate.	implementate doar la nivel individual (Awa, Plaumann şi Walter, 2010).
Diseminarea protocoalelor de intervenţie în practică Această componentă de cercetare, specifică categoriei cercetare-inovare-dezvoltare, vizează implementarea în activitatea practică a intervenţiilor care au trecut de filtrele anterioare. Dincolo de aspectul transferului de servicii psihologice inovative în practica profesională, în acest stadiu se pot avea în vedere studii cantitative sau calitative menite să identifice măsura în care beneficiarii acestor intervenţii le acceptă şi apreciază (complianţa la intervenţie), precum şi analize focalizate pe avantaje colaterale (cum ar fi un echilibru mai ridicat muncă-viaţă de familie) sau pe câştigul obţinut aşa cum rezultă în urma unei analize de cost-eficienţă.	În domeniul consultanţei manageriale (atât la nivelul companiilor de top ca Deloitte, Bain, BCG sau McKinsey, cât şi a firmelor mici de consultanţă) se pot oferi pachete de intervenţie menite să reducă nivelul de *burnout*. Ele pot avea ca suport diseminarea celor mai bune practici în domeniu, menite să reducă nivelul de epuizare emoţională în una sau mai multe arii precum motivare, autonomie, valori etc. prezentate în manuale de profil precum Maslach şi Leiter (1997) sau Leiter şi Maslach (2005). La fel de utile se pot dovedi şi programele generale de îmbunătăţire a sănătăţii mentale, cu efecte pozitive indirecte şi la nivel de *burnout*. De exemplu, programul de reducere a stresului bazat pe meditaţie *mindfulness* este aplicat în peste 200 de organizaţii din lumea întreagă cu scopul de a preveni *burnoutul* (Goutas, 2008)[18].

3.6. Întrebarea studiului – strategii de selectare a ideilor de cercetat

Lecturând literatura de specialitate pe o anumită tematică, ni s-a întâmplat poate multora dintre noi să ajungem să afirmăm că s-au scris atât de multe despre tema respectivă încât nu ar mai putea fi nimic nou de spus. Scopul

acestei secţiuni finale a capitolului destinat analizei critice a demersului de cercetare este de a oferi sugestii celor interesaţi cu privire la posibile idei de cercetat, inclusiv în acele domenii care par a fi deja bătătorite.

Desigur, o primă posibilitate este oferită de modelul pragmatic al ciclului ştiinţific natural, descris în secţiunea anterioară. În baza acestuia pot fi adresate o serie de întrebări precum: „cât de extinsă este problema (prevalenţa)?”, „care sunt factorii de risc antecedenţi?”, „există intervenţii eficiente pentru a preveni şi/sau combate problema?” etc.

O altă posibilă viziune este schematizată în figura 3.8, reprezentând o adaptare după Howitt şi Cramer (2000). Conform acestuia, una dintre sursele generative de idei de cercetare este analiza critică a cercetărilor existente pentru a evalua: (a) măsura în care afirmaţiile conţinute au acoperire empirică (sunt valide intern, extern, la nivel de construct şi statistic), propunând, acolo unde este cazul, noi studii cu metodologii adecvate, menite să corecteze eventualele aspecte rămase descoperite; (b) măsura în care rezultatele obţinute converg spre acelaşi rezultat sau conduc la rezultate mixte, contradictorii, propunând fie sinteze de tip metaanalitic pentru a surprinde efectul mediu global şi eventualii factori mediatori sau moderatori, fie studii empirice cu un design care să ofere explicaţii/sugestii cu privire la sursa efectelor contradictorii obţinute.

Tot din analiza literaturii de specialitate pot rezulta idei de cercetare provenite din: (a) sugestiile oferite de autorii studiilor cu privire la posibile direcţii viitoare de studiu; (b) testarea unor afirmaţii (conţinuturi teoretice) care nu au fost abordate empiric sau care au fost tratate într-o manieră insuficientă pentru a putea aprecia gradul de sprijin empiric existent pentru acestea.

În sfârşit, a treia categorie de surse de idei vizează o abordare diferită de cele anterioare. Dacă în cazurile precedente ideile de cercetare erau

bazate pe un fundament teoretic solid şi constituie cel mai adesea un demers logic de tip deductiv (dinspre teoretic spre empiric), idei de cercetare pot rezulta şi printr-un demers de tip inductiv, de la o realitate relevantă de cercetat spre o conceptualizare teoretică ulterioară. Aici ar putea intra (a) atât nevoia de a oferi răspunsuri la probleme noi cu care se confruntă societatea actuală (de exemplu, studierea eficienţei terapiilor realizate online, a impactului reţelelor de comunicare socială de tipul Facebook asupra vieţii de cuplu etc.), (b) cât şi ideile având drept sursă insighturi personale, precum şi problemele de zi cu zi cu care se confruntă cercetătorul (de exemplu, dacă cercetătorul observă din activitatea de consultanţă că persoanele agreabile primesc în general evaluări mai bune din partea supervizorilor decât cele mai puţin agreabile, îşi poate pune problema dacă agreabilitatea este într-adevăr un predictor al performanţei profesionale sau dacă relaţia evidenţiază o distorsiune în evaluare, aceea de a evalua mai intens pozitiv pe cei cu un stil de relaţionare interpersonală plăcut).

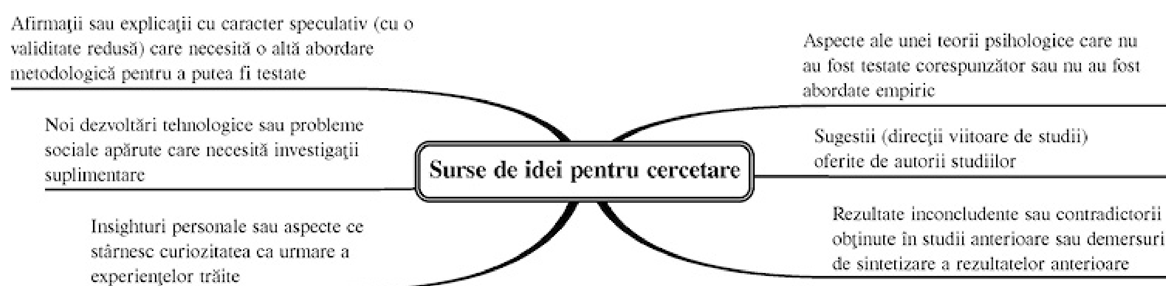


Figura 3.8. *Ciclul ştiinţific natural al studierii unui aspect psihologic – o viziune pragmatică*

Dată fiind importanţa acordată problemelor de validitate ale unei cercetări prezentate pe parcursul acestui capitol, voi oferi câteva detalii despre modul concret în care pot fi găsite idei de cercetare pornind de la analiza validităţii concluziilor unui studiu. Cel mai probabil, această sursă de idei are cel mai ridicat nivel de fertilitate, îndeosebi în contextul actual

de cercetare, în care suntem confruntaţi cu o explozie de studii care ne determină să credem că este dificil să aducem contribuţii originale, că lucrurile importante despre un subiect de cercetat au fost spuse deja. Lista acestor propuneri este prezentată în tabelul 3.13, fără a avea pretenţia de a fi o listă exhaustivă.

Tabelul 3.13. *Idei de studiu pornind de la analiza validităţii rezultatelor studiilor anterioare*

Tip de situaţie	Simptomatologie (ilustrare)/Soluţie (ilustrare)
Efect nedetectat	*Ipoteza studiului nu este sprijinită, nefiind obţinută o corelaţie semnificativă statistic între nivelul de conştiinciozitate şi performanţa în muncă.*
Validitate statistică	Posibile cauze pentru această stare de fapt sunt restricţia de amplitudine sau o consistenţă internă scăzută a variabilelor măsurate şi, în general, o putere statistică scăzută asociată ipotezei testate. *Soluţia*: de aplicat aceeaşi idee de cercetare pe un lot mai mare şi mai eterogen, cu instrumente având proprietăţi psihometrice superioare.
Validitate internă	Cu toate că aspectele de validitate internă au în vedere validarea afirmaţiilor de tip cauzal, am putea discuta şi în cazul acestui studiu corelaţional de probleme de instrumentare. De exemplu, dacă evaluarea performanţelor în muncă se face de către supervizori diferiţi, după bunul lor plac, fără a exista o metodologie comună de apreciere, se pot introduce erori de instrumentare care să conducă la rezultate nesemnificative statistic. *Soluţia*: preferabilă ar fi utilizarea unor indicatori hard (obiectivi) de apreciere a performanţei în muncă sau apelul la un training de profil şi la nişte ancore comportamentale de evaluare, dacă se apelează la evaluări soft (subiective).
Validitate de construct	O posibilă explicaţie ar fi utilizarea unui instrument mai puţin cunoscut (eventual cu probleme privind validitatea convergentă) pentru a măsura conştiinciozitatea. *Soluţia*: atât pentru conştiinciozitate, cât şi pentru aprecierea performanţelor ar trebui utilizate instrumente consacrate sau alte metode decât cele utilizate iniţial (de exemplu, măsuri obiective ale criteriului în locul evaluărilor oferite de supervizori); o altă direcţie posibilă ar viza descompunerea constructului conştiinciozitate pentru a urmări care dintre faţetele sale (de exemplu, prudenţa) corelează sau nu cu performanţa în muncă.

Validitate externă	Este posibil ca relaţia dintre conştiinciozitate şi performanţă în muncă să fie nesemnificativă pentru că s-a apelat la un lot de actori. *Soluţia*: se replică studiul apelând la un lot de sportivi de performanţă sau pe un lot de consultanţi financiari; se poate modifica contextul testării, aplicându-se proba în context de selecţie, nu într-unul tipic situat temporal la momentul postselecţiei.
Efect detectat	*Ipoteza studiului este susţinută, un nivel mai ridicat al anxietăţii sociale fiind înregistrat în cazul celor din grupul de control de tipul listei de aşteptare comparativ cu cei din grupul experimental, care au urmat trainingul de asertivitate.*
Validitate statistică	Cel mai probabil, acest tip de validitate nu ridică probleme în cazul descris cu două excepţii: (a) situaţia raportărilor selective (limităm discuţia la sfera anxietăţii sociale şi concluzionăm că trainingul este unul eficient, deşi nu au apărut diferenţe între cele două grupe cu privire la alte variabile studiate precum asertivitatea, empatia, stima de sine etc., singura diferenţă înregistrată fiind la nivelul anxietăţii sociale – care ar fi putut ieşi semnificativă statistic din întâmplare – eroarea setului de comparaţii); (b) situaţia puţin probabilă a derulării studiului cu o ipoteză de lucru căreia îi este asociată o putere statistică foarte mare. *Soluţia*: alegerea unei variabile dependente principale şi derularea studiului în condiţiile unei puteri statistice optime.
Validitate internă	Dacă rezultatul a fost obţinut în absenţa randomizării prealabile a participanţilor, rezultatul poate fi provocat de distorsiuni de selecţie. Dacă există randomizare, este posibil ca rezultatul obţinut să fie cauzat de alte aspecte precum pierderea selectivă a participanţilor (doar 55% dintre persoanele din lotul de control au fost disponibile pentru evaluare la nivel de postintervenţie comparativ cu 80% din grupul care a primit intervenţia). *Soluţia*: apelul la studii randomizate individual sau la nivel de cluster în locul celor de tip cvasiexperimental, cu derularea studiului în mai bune condiţii pentru a evita probleme tipice precum pierderea selectivă a participanţilor, istoricul etc.
Validitate de construct	Dacă rezultatul este unul pozitiv se pune întrebarea: care ar fi explicaţia pentru acest rezultat? Să fie simplul fapt că indivizii din grupul experimental au beneficiat de atenţie specială? Să fie din cauza faptului că trainingul, prin expunere repetată la situaţii sociale, scade nivelul de stres al participanţilor la studiu, astfel încât aceştia nu mai sunt la fel de anxioşi social? *Soluţia*: pentru a vedea dacă efectul se datorează trainingului propus şi nu altor factori nespecifici se apelează la un altfel de grup de

	control, de tipul intervenţiei placebo; pentru a oferi explicaţii privind efectul găsit se pot introduce măsurători ale posibilelor mecanisme ale schimbării, pentru a testa efectul acestora.
Validitate externă	Poate fi replicat rezultatul în alte contexte? De exemplu, pe subiecţi care au un nivel clinic al anxietăţii sociale? Sau în contextul în care trainingul este livrat colectiv, nu în şedinţe individuale? În ţări cu o cultură colectivistă faţă de cele care au o cultură individualistă? Etc. *Soluţia*: Implementarea programului de training în contexte cât mai variate şi pe participanţi cât mai diferiţi, pentru a urmări gradul de generalizare a rezultatelor.
Alte situaţii des întâlnite	
Efect stabil robust	Din multe studii reiese faptul că terapia cognitiv-comportamentală (TCC) este eficientă pentru a diminua simptomatologia depresivă (rezultate semnificative statistic). *Soluţia*: Dacă efectul a fost replicat în contexte variate se pot căuta explicaţii (factori mediatori) pentru legătura dintre VI şi VD. De exemplu, se poate testa măsura în care un anumit tip de cogniţii iraţionale, precum sunt cele autodepreciative, pot juca rolul unor variabile mediatoare, astfel încât terapia să modifice mai întâi nivelul acestor cogniţii, care la rândul lor au un impact asupra nivelului de depresie.
Efect variabil (eterogen)	Din studii diferite reies rezultate eterogene, astfel încât în unele cazuri se susţine empiric ideea că TCC are rol benefic în combaterea depresiei, iar în alte studii nu se replică acest rezultat pozitiv, nefiind înregistrate diferenţe semnificative statistic. *Soluţia*: introducerea în studiu a unor variabile moderatoare pentru a analiza situaţiile în care efectul este unul mai puternic, respectiv unul absent sau mai slab dezvoltat.
Replicarea conceptuală	Într-un studiu s-a demonstrat că vizionarea de filme agresive (grupul experimental) *vs* filme documentare (grupul de control) conduce la un nivel mai ridicat al agresivităţii de stare măsurată printr-o scală de tip *self-report*. Acest gen de rezultat este cu atât mai robust cu cât poate fi replicat prin apelul la alte modalităţi de operaţionalizare decât cele folosite iniţial. *Soluţia*: de exemplu, cei din grupul experimental sunt amorsaţi cu o serie de cuvinte cu sens din sfera agresivităţii (cuvinte cu o conotaţie agresivă), în timp ce participanţii din grupul de control au fost amorsaţi cu o serie de cuvinte neutre. VD a fost evaluată într-o manieră obiectivă, fiind înregistrate numărul de conflicte (agresiuni verbale) într-o sarcină de lucru primită.

Concluzii

În acest capitol au fost discutate principalele elemente necesare pentru a înţelege dimensiunea cercetării ştiinţifice în psihologie. Cu toate că sunt precizate aspecte specifice pentru metodologiile calitative, accentul este pus pe studiile care utilizează o metodologie cantitativă, acesta fiind curentul majoritar în domeniul psihologiei.

Însă indiferent de tipul de cercetare avut în vedere, aprecierea calităţii acesteia nu se face luând în seamă criterii cantitative superficiale, de genul numărului de ipoteze, a numărului de instrumente utilizate, a complexităţii analizelor statistice etc., ci de măsura în care studiile respective sunt valide. Validitatea unui studiu nu este determinată *per se*, ci are în vedere nuanţe diferite în funcţie de obiectivele sale şi nici nu priveşte studiul în ansamblul său, ci se analizează la nivelul afirmaţiilor (concluziilor) din studiu. Astfel un rezultat statistic poate fi valid (validitate statistică), fără a fi însă adevărată relaţia cauzală specificată între variabilele studiate (validitate internă scăzută). La fel, un rezultat poate fi, atât valid statistic, cât şi clar din perspectiva unei relaţii cauzale, însă explicaţia oferită pentru această stare de fapt să fie una greşită (validitate de construct scăzută). În sfârşit, chiar dacă rezultatul este semnificativ statistic şi indică o relaţie cauzală bazată pe o explicaţie corect identificată, este posibil ca efectul observat să aibă un grad scăzut de generalizare (transfer) în alte contexte (validitate externă scăzută).

Aşadar, în analiza validităţii trebuie urmărite toate cele patru forme ale sale, îndeosebi pe acelea cu cel mai ridicat nivel de relevanţă pentru un anumit tip de studiu. În acest sens, studiile pot viza obiective exploratorii, obiective descriptive, obiective cauzale şi obiective explicative (cauzal-explicative). În funcţie de obiectivul vizat de un

anumit studiu pot fi deosebit de importante anumite tipuri de validitate: (a) de construct (pentru obiective exploratorii); de construct, externă şi statistică (pentru obiective descriptive); internă (pentru obiectivele cauzale); internă şi de construct (pentru obiective cauzal-explicative). În sfârşit, din perspectiva designului utilizat, studiile cu un nivel ridicat de rigurozitate, precum sunt experimentele, necesită o atenţie deosebită aspectelor ce ţin de validitatea internă şi a celei de construct. Cunoaşterea în profunzime a aspectelor legate de validitatea unei cercetări este importantă nu numai din perspectivă critică, ci şi pentru a identifica idei viitoare de cercetare.

În sfârşit, deoarece niciun studiu nu poate răspunde perfect tuturor riscurilor posibile pentru toate formele de validitate, se recomandă la nivelul unui studiu individual centrarea pe acele aspecte ale validităţii importante pentru a răspunde adecvat întrebării studiului. Numai programele de cercetare pe o anumită temă pot răspunde adecvat şi în timp (pe rând), la fiecare dintre posibilele limite sesizate. Pe baza mai multor asemenea programe de studii se poate ajunge la completarea unui ciclu întreg de studii cu referire la problematica investigată. În cadrul capitolului am prezentat ciclul natural ştiinţific de abordare a unei probleme într-o manieră pragmatică. Acesta poate fi descris succint prin două mari faze: (a) faza de cunoaştere a problemei şi (b) faza de soluţionare a problemei.

Lecturi suplimentare

Alan E. Kazdin (2010). *Research design in clinical psychology* (ediţia a IV-a). Boston, USA: Allyn & Bacon.

Capitolele 2, 3 şi 4 ale cărţii tratează pe larg problematica celor patru tipuri de validitate şi a distorsiunilor legate de activitatea de cercetare în general. Referinţa este îndeosebi utilă celor interesaţi de domeniul clinic, prin ilustrările pe care le conţine, însă ideile şi principiile enunţate sunt valabile pentru orice domeniu de cercetare psihologică.

William R. Shadish, Thomas D. Cook, Donald T. Campbell (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, USA: Houghton Mifflin Company.

Cartea reprezintă într-o manieră actualizată informaţiile prezentate în volumele clasice din domeniu cum ar fi Cook şi Campbell (1979), tratând într-o manieră extensivă problematica validităţii unei cercetări din perspectiva afirmaţiilor pe care le incumbă.

John W. Creswell (2007). Qualitative inquiry & Research design. Choosing among five approaches. Londra, UK: Sage Publications.

Recomand această lucrare pentru cei interesaţi să îşi dezvolte abilităţile de cercetare calitativă. Pe parcursul lucrării sunt prezentate pe larg cinci dintre cele mai cunoscute tipuri de cercetare calitativă: abordarea narativă (cum ar fi istoria orală), fenomenologia, cercetările etnografice, teoria generată (*grounded theory*) şi studiul de caz intensiv.

Podsakoff, P.M., MacKenzie, S.M., Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendation on how to control it. *Annual Review of Psychology*, 63, pp. 539-569.

Articolul prezintă impactul pe care îl exercită distosiunea metodei comune, probabil cea mai întâlnită problemă legată de validitatea de construct, asupra rezultatelor statistice obţinute, precum şi o serie de măsuri, menite să reducă sau să controleze nivelul acestei distorsiuni.

Bibliografie

Aronson, E., Mills, J. (1959).The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 59, pp. 177-181.

Awa, W., Plaumann, M., Walter, U. (2010). Burnout prevention: A review of intervention programs. *Patient Education and Counseling*, 78, pp. 184-190.

Bedos. C., Pluye. P., Loignon, C., Levine, A. (2008). Qualitative research. În E. Lesaffre, J. Feine, B. Leroux, D. Declerck (eds.), *Statistical and methodological aspects of oral health research* (pp. 113-130). West Sussex, UK: John Wiley and Sons.

Bontilă, G. (1935). *Tehnica testării psihologice*. Bucureşti: Editura Cartea Românească.

Brewer, M.B. (2000). Research design and issues of validity. În H.T. Reis, C.M. Judd (eds.), *Handbook of research methods in social and personality psychology* (pp. 3-16). New York: Cambridge University Press.

Campbell, D.T., Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. În N.L. Gage (ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.

Carroll, K.M., Nich, C., Sifry, R.L., Nuro, K.F., Frankforter, T.L., Ball, S.A. *et al.* (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and alcohol dependence*, 57, pp. 225-238.

Cook, T.D. Campbell, D.T.(1979). *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally.

Cordray, D.S., Pion, G.M. (2006). Treatment strength and integrity: Models and methods. În R.R. Bootzin, P.E., McKnight (eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103-124). Washington, DC: American Psychological Association.

Creswell, J.W. (2007). *Qualitative Inquiry & Research Design. Choosing among five approaches*. Londra: Sage Publications.

Creswell, J.W. (2006). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (ediţia a III-a). Upper Saddle River, NJ: Pearson.

David, D. (2006). *Metodologia cercetării clinice*. *Fundamente.* Iaşi: Editura Polirom.

David, D., Szentagotai, A., Lupu, V., Cosman, D. (2008). Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: A randomized clinical trial, post-treatment outcomes, and six-month follow-up. *Journal of Clinical Psychology*, 64, pp. 728-746.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, 10, pp. 230-241.

Demerouti, E., Bakker, A.B., Nachreiner, F., Schaufeli, W.B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, 86, pp. 499-512.

Elliott, R., Timulak, L. (2005). Descriptive and interpretative approaches to qualitative research. În J. Miles, P. Gilber, (eds.), *A Handbook of Research Methods for Clinical and Health Psychology* (pp.147-160). Marea Britanie: Oxford University Press.

Elmes, D.G., Kantowitz, B.H., Roediger, H.L. (2012). *Research methods in psychology* (ediţia a IX-a). St. Paul, MN: West.

Freudenberger, H.J. (1975). The staff burnout syndrome in alternative institutions. *Psychotherapy: Theory, Research and Practice*, 12, pp. 73-82.

Gawronski, B., Payne, B.K. (eds.) (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford Press.

Gilbert, P., Irons, C. (2008). Thinking about research: issues and dilemmas. În J. Miles, P. Gilber, (eds.), *A Handbook of Research Methods for Clinical and Health Psychology* (pp.1-14). Marea Britanie: Oxford University Press.

Goutas, L.J. (2008). Bournout. *The Write Stuff*, 17, pp. 135-138.

Greenwald, A.G., McGhee, D.E., Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, pp. 1464-1480.

Heppner, P.P., Wampold, B.E., Kivlighan, D.M., Jr. (2008). *Research design in counseling* (ediţia a III-a). Belmont, CA: Thomson Brooks/Cole.

Howitt, D., Cramer, D. (2000). *First steps in research and statistics: A practical workbook for psychology students*. Londra: Routledge.

Kantowitz, B.H., Roediger, H.L., III, Elmes, D.G. (2009). *Experimental Psychology*. Cheriton House, UK:Cengage Learning.

Kaschka, W.P., Korczak, D., Broich, K. (2011). Burnout: a fashionable diagnosis. *Deutsches Ärzteblatt*, 108, pp. 781-787. doi:10.3238/arztebl.2011.0781

Kazdin, A.E. (2010). *Single-case Research Designs: Methods for Clinical and Applied Settings* (ediţia a II-a). New York: Oxford University Press.

Kerlinger, F.N., Lee, H.B. (2000). *Foundations of behavioral research* (ediţia a IV-a). Holt, NY: Harcourt College Publishers.

Kraemer, H.C. (2000). Pitfalls of multisite randomized clinical trials of efficacy and effectiveness. *Schizophrenia Bulletin*, 26, pp. 533-541.

Laverty, S.M. (2003). Hermeneutic phenomenology and phenomenology: A comparison of historical and methodological considerations. *International Journal of Qualitative Methods*, 2(3). Găsit la adresa http://www.ualberta.ca/~iiqm/backissues/2_3final/pdf/laverty.pdf

Leiter, M.P., Maslach, C. (2005). *Banishing burnout: Six strategies for improving your relationship with work*. San Francisco, CA: Jossey-Bass.

LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre, J., Derderian, F. (1997). Discrepancies between metaanalyses and subsequent large randomized, controlled trials. *The New England Journal of Medicine*, 337, pp. 536-542. doi:10.1056/NEJM199708213370806

Maslach, C. (1976). Burned-out. *Human Behavior*, 9, pp. 16-22.

Maslach, C. (1981). Burnout: A social psychological analysis. În J.W. Jones (ed.), *The burnout syndrome* (pp. 30-53). Park Ridge, IL: London House.

Maslach, C., Leiter, M.P. (1997). *The truth about burnout*. San Francisco, CA: Jossey-Bass.

Maslach, C., Schaufeli, W.B., Leiter, M.P. (2001). Job burnout. *Annual Review of Psychology*, 52, pp. 397-422. doi:10.1146/annurev.psych.52.1.397

Maxwell, J.A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62, pp. 279-301.

Olantunji, B.O., Lohr, J.M., Sawchuk, C.N., Westendorf, D.H. (2005). Using facial expressions as CSs and fearsome and disgusting pictures as UCSs: affective responding and evaluative learning of fear and disgust in blood-injection-injury phobia. *Journal of Anxiety Disorders*, 19, pp. 539-555.

Pelham, B.W., Blanton, H. (2007). *Conducting research in psychology: Measuring the weight of smok*e (ediţia a III-a). Belmont, CA: Wadsworth/Thomson.

Podsakoff, P.M., MacKenzie, S.M., Lee, J., Podsakoff, N.P. (2003). Common method variance in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, pp. 879-903.

Podsakoff, P.M., MacKenzie, S.M., Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendation on how to control it. *Annual Review of Psychology*, 63, pp. 539-569.

Raudenbush, S.W., Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5, pp. 199-213.

Sava, F. (2004/2011). *Analiza datelor în cercetarea psihologică. Metode statistice complementare.* Cluj-Napoca: ASCR.

Sava, F.A. (ed.) (2011). *Metode implicite de investigare a personalităţii*. Iaşi: Editura Polirom.

Schaufeli, W.B. (2003). Past performance and future perspectives of burnout research. *South African Journal of Industrial Psychology*, 29, 1-15.

Schaufeli, W.B., Leiter, M.P., Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International*, 14, pp. 204-220. doi:10.1108/13620430910966406

Scifres, C.M., Iams, J.D., Klebanoff, M., Macones, G.A. (2009). Metaanalysis vs large clinical trials: which should guide our management?. *American Journal of Obstetrics & Gynecology*, 5, p. 484.

Shadish, W.R., Cook, T.D., Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Singera, R.A., Zentall, T.R. (2011). Preference for the outcome that follows a relative aversive event: Contrast or delay reduction? *Learning and Motivation*, 42, pp. 255-271.

Susser, E., Schwartz, S., Morabia, A., Bromet, E. (2006). *Psychiatric epidemiology: Searching for the causes of mental disorders*. New Zealand: Oxford University Press.

Steckler, A., McLeroy, K.R., Goodman, R.M., Bird, S.T., McCormick, L. (1992). Toward integrating qualitative and quantitative methods: An introduction. *Health Education Quarterly*, 19, pp. 1-8.

Sweldens, S., van Osselaer, S.M.J., Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of Consumer Research*, 37, pp. 473-489. doi:10.1086/653656

Tashakkori, A., Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Applied Social Research Methods, Nr. 46). Thousand Oaks, CA: Sage.

Turok, K.D., Espez, E., Edelman, A.B., Lotke, P.S., Lathrop, E.H., Teal, S.B., Jacobson, J.C., Simonsen, S.E., Schulz, K.F. (2011). The methodology for developing a prospective metaanalysis in the family planning community. *Trials*, 12, p. 104. doi:10.1186/1745-6215-12-104

Trochim, W.M. (2006). *The Research Methods Knowledge Base* (ediţia a II-a). Găsit la 10.08.2012 la adresa http://www.socialresearchmethods.net/kb/index.php.

van der Klink, J.J, Blonk, R.W., Schene, A.H., van Dijk, F.J. (2001). The benefits of interventions for work-related stress. *American Journal of Public Health*, 91, pp. 270-276.

Whittemore, R., Chase, S.K., Mandle, C.L. (2001). Validity in qualitative research. *Qualitative Health Research*, 11, pp. 522-537.

Wikipedia (2012). *Effort justification*. Găsit la data de 05.05.2012 la adresa: http://en.wikipedia.org/wiki/ Effort_justification.

1. În ţările dezvoltate, orice proiect de cercetare poate fi implementat doar după ce primeşte aprobarea din partea unei comisii de etică a cercetării (*IRB – Institutional Review Board*). O asemenea practică a început să fie implementată recent şi la noi, vezi cazul Universităţii „Babeş-Bolyai” din Cluj-Napoca. În plus, pentru a putea derula cercetări pe subiecţi umani, unele instituţii solicită cercetătorilor să obţină/să facă dovada existenţei unui certificat de absolvire cu succes a unui curs despre protejarea participanţilor la cercetare. Un exemplu de astfel de curs în limba engleză, ce poate fi urmat gratuit de oricine, este disponibil la http://phrp.nihtraining.com/users/overview.php

2. În acest sens, în special în domeniul ştiinţelor medicale se vorbeşte tot mai des despre cercetarea translaţională ca un tip de cercetare liant între cea fundamentală şi cea aplicată, având rolul de a grăbi transferul rezultatelor obţinute în cercetarea fundamentală (de exemplu, anumite descoperiri din biologia moleculară) în zona clinică (a studiilor randomizate), precum şi de a asigura feedbackul invers, astfel încât anumite probleme stringente în zona clinică să determine agenda cercetării din zona fundamentală. Acest tip de cercetare necesită expertiză avansată şi o abordare interdisciplinară.

3. Pentru a ilustra şi mai clar consecinţele aplicării unei metodologii inadecvate vom apela la exemplul oferit de David (2006) cu privire la Moş Crăciun. Astfel, dintr-o cercetare efectuată în paradigma hermeneutică cu privire la portretul psihologic al lui Moş Crăciun la copiii de 4-5 ani am concluziona corect (fiind vorba de credinţele lor) că: (a) există probe despre existenţa lui Moş Crăciun; (b) dacă îl vom îmbuna, fiind cuminţi, vom primi mai multe cadouri. Cu alte cuvinte, rezultatele studiului nu arată că aşa stau lucrurile (că Moş Crăciun există), ci că copiii cred că Moş Crăciun există. Printr-un demers pozitivist, am afla însă că Moş Crăciun nu există (ce realitate dură!).

Acest exemplu poate părea ilar, pentru că toţi cititorii acestei cărţi cunosc această diferenţă, însă dacă ne referim la aspecte psihologice complexe supuse cercetării, vom observa cazuri în care cercetătorii confundă realitatea observată cu realitatea construită (luând credinţele participanţilor drept adevăruri pentru a susţine, de pildă, că plecarea părinţilor la muncă în străinătate cauzează tulburări emoţionale copiilor rămaşi acasă). Pericolul acestui tip de confuzie este cu atât mai mare, cu cât credinţele împărtăşite de participanţi par a fi unele rezonabile logic (de exemplu, că programele educaţionale cu privire la conduita în trafic adresate adolescenţilor vor reduce numărul accidentelor în care sunt implicaţi, afirmaţie nesusţinută empiric – vezi exemplul al doilea din caseta 2.1, capitolul 2). Cu toate acestea,

cercetarea credinţelor (a realităţii construite) este la fel de onorabilă şi utilă precum demersurile de tip pozitivist. Astfel, pentru activitatea unui consilier şcolar ar fi foarte important să se cunoască modul în care adolescenţii din familii cu părinţi aflaţi la muncă în străinătate, percep şi trăiesc această stare de fapt, respectiv modul în care tinerii se raportează la programele de educaţie rutieră, pentru a le putea îmbunătăţi.

4. Distincţiile prezentate în acest tabel au fost făcute în scop pedagogic, pentru a sublinia unele diferenţe între cele două abordări. În realitate între cele două viziuni există mai degrabă diferenţe de nuanţă. De exemplu, abordarea pozitivistă nu susţine că nu există cunoaştere (realitate) subiectivă, ci doar că analiza acesteia duce la studierea unei realităţi percepute, nu a realităţii obiective.

5. Există o distincţie clară între conceptele de metodologie a cercetării, designul cercetării şi metodele de cercetare, relaţia dintre ele fiind una ierarhică. În funcţie de metodologia aleasă (de exemplu, o metodologie cantitativă în paradigma pozitivistă), se apelează la un anumit design de cercetare (de exemplu, unul de tip descriptiv, non-experimental, sub forma unui sondaj de opinie), care, la rândul său, presupune utilizarea unor metode de cercetare adecvate (de exemplu, metoda chestionarului). Astfel, Creswell (2006) sintetizează distincţia conceptuală dintre cele trei. Prin metodologia cercetării se stabilesc liniile mari prin care se va derula o cercetare ca urmare a unei opţiuni filosofice pentru o paradigmă de cercetare (un mod de a ne raporta şi înţelege lumea şi de a acumula cunoaştere ştiinţifică, fiind lentilele care ne direcţionează spre ce anume privim în lume). Prin designul cercetării se stabileşte un plan de acţiune prin care se încearcă să fie valorificate metodele de cercetare pentru a atinge obiectivele studiului, fiind astfel un nivel intermediar de raportare între aspectele filosofice (din sfera metodologiei cercetării) şi aspectele tehnice (din sfera metodelor de cercetare). Prin metode de cercetare ne referim la tehnici specifice de colectare a datelor, cu implicaţii la nivelul analizei datelor (interviu, observaţie participativă, teste psihologice etc.).

6. Deşi interrelaţionate, nu există o echivalenţă între obiectivele descriptive şi metodele de cercetare utilizate. De exemplu, sondajele de opinie pot viza aspecte analitice, în timp ce studiul bazat pe observaţii poate viza obiective de tip exploratoriu. În mod similar, studiile de caz (menţionate ca studii exploratorii) pot viza obiective descriptive.

7. De exemplu, în studiile experimentale întotdeauna manipularea VI precedă măsurarea efectelor la nivel de VD cauzate de această manipulare. De asemenea, în studiile non-experimentale longitudinale, se pot evidenţia legături între factorul cauzal potenţial măsurat la momentul T0 şi consecinţele asupra variabilei efect la momentul T1.

8. În unele cazuri, legătura statistică dintre variabile poate fi unidirecţională din punct de vedere logic. Un exemplu ar fi asocierea dintre temperatura ambientală ridicată şi un nivel crescut de ostilitate (iritabilitate). Ar fi dificil logic să gândim că pe măsură ce oamenii devin mai iritaţi creşte temperatura mediului ambient, în schimb ar fi logic plauzibil să afirmăm că pe măsură ce creşte temperatura din mediul ambient oamenii devin mai iritaţi (ostili).

9. Pe lângă riscul în sine, constituie pericole la adresa validităţii interne a unui studiu şi interacţiunile dintre distorsiunea de selecţie şi alte riscuri (de exemplu, distorsiunea de selecţie × istoric; distorsiunea de selecţie × maturare etc.). De exemplu, este posibil ca persoanele incluse în grupul experimental să fie mai preocupate de aspectul durerii şi să caute să se informeze în legătură cu tehnici psihologice de reducere a durerii. În acest caz ipotetic, cele două grupe vor diferi nu numai în ceea ce priveşte deprinderea tehnicii Lamaze, ci şi în privinţa nivelului de cunoaştere a altor tehnici (recomandări psihologice) menite să reducă durerea.

10. Desigur, ideal ar fi să se tindă către standardul maxim aplicabil în domeniul ştiinţelor medicale, îndeosebi al farmacoterapiei, în care studiul experimental cu grup de control activ (intervenţie standard) se derulează în condiţii lipsite de transparenţă (cum ar fi designul triplu orb, în care nici participanţii la studiu, nici cei care pun la dispoziţia participanţilor medicaţia analizată în studiu şi nici asistenţii de cercetare care evaluează efectele obţinute nu ştiu cărei condiţii experimentale îi aparţine cel evaluat). Într-un astfel de design sunt controlate eficient toate riscurile posibile privitoare la validitatea internă a studiului. Din păcate, în cazul intervenţiilor psihologice este imposibil de implementat un asemenea studiu, cel puţin cu privire la situaţia celor care livrează intervenţia. Totuşi, pot fi luate măsuri în direcţia reducerii transparenţei obiectivului studiului, fie prin aplicarea individuală, fie prin combinarea unor măsuri precum: (a) necomunicarea ipotezei urmărite celor care livrează, respectiv primesc intervenţia; (b) utilizarea unor asistenţi de cercetare pentru colectarea măsurătorilor de tip pretest, interim sau posttest care să nu cunoască nici cărei condiţii experimentale îi aparţine cel examinat şi/sau nici care ar fi ipoteza studiului; (c) utilizarea unor intervenţii nespecifice, credibile pentru participanţii la studiu, pentru a decela efectul placebo de efectul real al intervenţiei.

11. Pelham şi Blanton (2007) menţionează o altă distincţie posibilă între validitatea de construct şi validitatea conceptuală. Prima vizează mai degrabă legătura dintre concept (construct) şi modul de operaţionalizare al său în cadrul unui studiu (de exemplu, reprezintă seria de desene *Ben 10* o modalitate adecvată de operaţionalizare pentru conceptul de desene animate agresive?). În schimb, validitatea conceptuală are un sens mult mai larg şi se referă la măsura în care întregul concept studiat poate fi explicat din perspectiva unei teorii, la măsura în care un întreg set de studii (program de studii) sprijină sau nu o anumită teorie, utilitatea unui anumit concept nou.

12. Un alt posibil risc clasic, rar întâlnit la nivelul studiilor publicate în reviste cu prestigiu internaţional, este cel al definirii inadecvate a conceptului studiat, cu consecinţe negative asupra operaţionalizării acestuia. În schimb, mai frecvent pot apărea probleme de etichetare a constructului studiat. De exemplu, într-un studiu despre relaţia dintre agresivitate şi consumul de filme video violente se pot înţelege prin agresivitate (inclusiv la nivel de operaţionalizare) numai aspecte ce se referă la agresivitatea fizică, fără a socoti alte faţete ale agresivităţii precum agresivitatea verbală sau atitudinea ostilă/pasiv-agresivă.

13. Reactivitatea participanţilor este trecută clasic ca un risc privitor la validitatea de construct (Shadish et al., 2002; Cook şi Campbell, 1979; Kazdin, 2010). De exemplu, atunci când

obţinem un rezultat semnificativ statistic şi vrem să explicăm rezultatul (scop cauzal-explicativ), nu am şti dacă efectul este cauzat de intervenţie (manipulare experimentală) sau este o consecinţă a reactivităţii participanţilor. În sens utilitar însă („este intervenţia eficientă?" „există o relaţie cauză-efect?"), reactivitatea participanţilor constituie un risc la nivelul validităţii interne, fiind un artefact ce creşte probabilitatea apariţiei unei erori de tip I. De exemplu, intervenţia poate să nu fie eficientă în realitate, dar rezultatul obţinut să fie unul semnificativ statistic, din cauza reactivităţii participanţilor.

14. Un ghid succint, dar foarte util despre debriefing găsiţi la: http://www.wadsworth.com/psychology_d/templates/student_resources/workshops/res_methd/debrief/debrief_1.html

15. Brewer (2000) semnalează că, într-un studiu bazat pe ameninţarea stimei de sine prin oferirea unui feedback negativ participanţilor pentru un eseu scris, în condiţii controlate de laborator, sarcina are un realism psihologic mai ridicat pentru subiecţi dacă aceştia sunt studenţi decât pentru mineri cu vârsta de peste 50 de ani.

16. Cel mai cunoscut cost de acest gen priveşte relaţia dintre validitatea internă şi cea externă. Pe măsură ce există mai multă preocupare faţă de controlul variabilelor confundate pentru a creşte validitatea internă, creşte totodată nivelul de artificialitate al cercetării, scăzând astfel validitatea externă. O altă relaţie inversă poate fi sesizată între validitatea de construct şi cea internă. De exemplu, pentru a evita distorsiuni precum subreprezentarea constructelor sau distorsiunea monometodă, cercetătorul poate apela la măsuri multiple de evaluare a efectului măsurat. Acest efect creşte însă timpul de examinare şi intensifică efortul participanţilor în cadrul studiului, fapt care poate conduce la creşterea procentului de participanţi care renunţă la studiu, cu consecinţe negative asupra validităţii interne (diferenţele dintre grupe sunt cauzate de retragerea diferenţiată din studiu) şi a celei externe (prin scăderea reprezentativităţii participanţilor rămaşi în studiu pentru populaţia generală).

17. Cititorii interesaţi de acest aspect al validităţii studiilor calitative pot găsi mai multe detalii, inclusiv despre aspecte specifice ale validităţii pentru diferite tipuri de studii (abordări fenomenologice, abordări narative etc.) în Creswell (2007).

18. Un asemenea ciclu de abordare ştiinţifică a unei probleme poate fi un instrument euristic util în a identifica punctele nevralgice ale unui domeniu de cercetare. În cazul specific al burnoutului, existenţa deopotrivă a unor intervenţii specifice, precum şi a celor general valabile pentru îmbunătăţirea stării de sănătate mentală poate fi interpretat ca o consecinţă directă a presiunii venite din lumea practicienilor de a oferi soluţii la probleme în contextul în care nu există studii elocvente cu privire la ingredientele active (mecanismele cauzale) responsabile pentru reducerea nivelului de burnout. Ca urmare a acestei stări de fapt, în care balanţa dintre cercetare fundamentală *vs* cercetare aplicativă înclină evident în favoarea ultimei, se impune sprijinirea mai multor cercetări fundamentale. Din acest punct de vedere, ar fi nevoie îndeosebi de o focalizare mai mare asupra obţinerii de dovezi empirice asupra modelelor teoretice dezvoltate pentru a explica burnoutul şi a identificării acelor mecanisme cauzale responsabile

pentru prevenirea şi/sau combaterea acestui fenomen nedorit, dar prezent în organizaţiile contemporane. Numai în acest fel vom putea să ne asigurăm că avem de-a face cu o abordare pragmatică, validată ştiinţific, iar practicienii nu vor mai putea fi acuzaţi că pun căruţa înaintea cailor, cu alte cuvinte că oferă soluţii pentru o problemă al cărei model teoretic este încă insuficient fundamentat empiric.

Partea a doua

Tipuri majore de studii în cercetarea psihologică

Capitolul 4

Studiile experimentale

4.1. Experimentul – metoda regală în cercetarea ştiinţifică

La cele mai multe dintre problemele menţionate pe parcursul primelor două capitole ale acestei lucrări, răspunsurile pot fi oferite prin apel la studii experimentale. Spre exemplu, putem analiza credinţa populară conform căreia ascultarea muzicii clasice compusă de Mozart are efecte benefice asupra dezvoltării intelectuale la copii. Mai specific, se consideră că ascultarea repetată a acestui tip de compoziţii conduce la potenţarea nivelului de inteligenţă al copiilor care sunt expuşi în mod repetat la această stimulare. Studiile non-experimentale, cum ar fi cele de tip corelaţional, nu sunt suficiente pentru a putea susţine ştiinţific o asemenea concluzie, deoarece este posibil ca simpla corelaţie dintre ascultarea muzicii lui Mozart şi coeficientul de inteligenţă a copiilor să fie doar un artefact statistic. În spatele ascultării muzicii lui Mozart s-ar putea regăsi situaţia unor familii cu un capital cultural bogat, fapt care facilitează dezvoltarea cognitivă a copilului. Singura metodă ştiinţifică prin care se poate demonstra existenţa sau inexistenţa efectului Mozart (şi, în general, eficienţa oricărui tip de intervenţie/tratament) este metoda experimentală.

Caseta 4.1. *Descrierea dovezilor ştiinţifice legate de efectul Mozart*

Rauscher, Shaw şi Ky (1993) au publicat rezultatele unui experiment în prestigioasa revistă *Nature*, prin care se susţinea ideea că muzica lui Mozart conduce la creşterea tranzitorie a performanţei în sarcini specifice pentru a surprinde inteligenţa spaţială. În sprijinul acestei concluzii, autorii au prezentat rezultatele unui experiment în cadrul căruia au existat trei grupe de participanţi: cei care au ascultat timp de 10 minute muzica lui Mozart (mai exact sonata pentru două piane K448), cei care au beneficiat timp de 10 minute de un instructaj specific tehnicilor de relaxare şi cei din condiţia neutră, care au stat timp de 10 minute în condiţii de linişte deplină. După manipularea variabilei de interes (ascultă sau nu muzica lui Mozart) şi controlul variabilelor potenţial confundate (prin distribuirea randomizată a participanţilor în cele trei grupe experimentale) cercetătorii au evaluat efectul muzicii prin măsurarea performanţei participanţilor în rezolvarea unor sarcini spaţiale din cadrul unui test de inteligenţă (componenta de inteligenţă spaţială). Rezultatele obţinute au uimit lumea ştiinţifică prin faptul că cei care au fost incluşi în grupul celor care au ascultat muzica lui Mozart au obţinut performanţe semnificativ statistic mai bune în sarcinile spaţiale decât cei din celelalte două grupe experimentale, diferenţa fiind de aproximativ 8-9 puncte în termeni de IQ, însă efectul erau unul tranzitoriu (se menţinea timp de 10-15 minute de la finalizarea expunerii la muzică). Studiile ulterioare au confirmat doar parţial aceste rezultate în sensul că cea mai recentă şi comprehensivă metaanaliză pe această temă, efectuată de Pietschnig, Voracek şi Formann (2010), indică un efect de intensitate moderată spre scăzută (d Cohen egal cu 0,37) a muzicii lui Mozart comparativ cu performanţa obţinută de cei care aşteptau în linişte (grupul de control), un efect similar fiind însă întâlnit şi dacă erau comparate alte compoziţii muzicale cu performanţele celor din grupul de control. Dacă se compara muzica lui Mozart cu alte genuri muzicale, diferenţa era una neglijabilă. Mai mult efectele observate sunt într-adevăr tranzitorii, sunt obţinute doar la nivelul componentei spaţiale a IQ şi se bazează îndeosebi pe studii făcute pe populaţie adultă.

Aceste rezultate nu au afectat însă dezvoltarea unei adevărate industrii de promovare a efectului Mozart pentru copii, susţinută de dorinţa părinţilor de a augmenta nivelul intelectual al copiilor. De exemplu, celebrul site Amazon.com conţinea în 2008 peste 40 de produse (CD-uri) care susţineau că sunt eficiente în producerea efectului Mozart, iar unele state americane au mers până într-acolo încât au votat un proiect de lege care cerea creşelor de stat să pună, în fiecare zi, muzică clasică sugarilor (Florida) sau au creat subvenţii din bugetul statului astfel încât fiecare copil nou-născut să primească gratuit un CD cu muzica lui Mozart (Georgia şi Tennessee) (Lilienfeld *et al.*, 2010).

Orice experiment presupune: (a) manipularea variabilei independente pentru a vedea efectele acestei acţiuni asupra variabilei dependente; (b) controlul variabilelor potenţial confundate în designuri intergrup prin alocarea randomizată a participanţilor în grupele experimentale. Când se vorbeşte despre ideea de control în cadrul unui experiment, se pot avea în vedere cel puţin trei niveluri diferite de analiză (Elmes, Kantowitz şi Roediger III, 2012): (a) nivelul comparativ (prin compararea performanţei grupului experimental cu cea a unui grup de control); (b) nivelul manipulării variabilei independente (VI) (control exercitat prin posibilitatea de a modifica nivelul intensităţii VI/tipul de intervenţie oferită participanţilor la studiu); (c) controlul variabilelor străine (potenţial confundate) care ar putea afecta relaţia de tip cauză-efect studiată (de exemplu, control exercitat prin alocarea randomizată a participanţilor în grupul experimental, respectiv în cel de control).

Prin manipularea variabilei independente (cauza) în cadrul unui experiment, cercetătorul exercită un control asupra a ceea *ce oferă* indivizilor (*input*), pentru a observa dacă apar modificări în realitate cu privire la efectul studiat (*output*). De exemplu, se poate urmări dacă performanţa indivizilor la sarcini vizual-spaţiale se modifică în contextul în care variabila-cauză este activă (se ascultă muzică de Mozart) sau inactivă (participanţii aşteaptă în linişte). În plus, poate fi testat aspectul de specificitate (doar muzica lui Mozart, orice tip de muzică clasică sau muzica, în general, produce aceste efecte).

De asemenea, prin alocarea randomizată a participanţilor în grupele experimentale, cercetătorul exercită un control asupra variabilelor potenţial confundate ce pot afecta rezultatele studiului. Este vorba de controlul acelor variabile care pot influenţa efectul urmărit – performanţa la sarcini spaţiale. Acest control se obţine prin egalizarea stării de fapt

existente în grupele experimentale cu referire la variabile precum atitudinea faţă de muzica clasică, dispoziţia afectivă, nivelul de cultură al participanţilor, nivelul IQ-ului lor general etc., aspect obţinut prin alocarea randomizată a participanţilor în grupele experimentale.

Dată fiind relevanţa definitorie pentru orice studiu experimental a conceptului de relaţie cauză-efect, precum şi a mijloacelor utilizate pentru atingerea acestui scop – manipularea variabilei cauză (independente), respectiv controlul variabilelor străine, vom prezenta, în cele ce urmează, fiecare dintre aceste concepte fundamentale pentru studiile experimentale.

4.1.1. Despre relaţia cauză-efect în psihologie

De obicei, cei mai mulţi dintre noi avem o predilecţie de a gândi în termenii unei relaţii de tip cauză-efect. Ne-am îmbolnăvit pentru că organismul nostru era slăbit ca urmare a lipsei de odihnă, avem probleme de somn pentru că suntem prea stresaţi, am promovat un examen dificil ca urmare a efortului depus de a învăţa la disciplina respectivă, copilul vecinului este neastâmpărat pentru că părinţii acestuia nu se ocupă suficient de educaţia lui etc. De multe ori, acest tip de argumentare este considerată a fi suficientă în gândirea uzuală pentru a concluziona asupra cauzelor respective că sunt suficiente pentru apariţia unui anumit eveniment.

Într-o abordare ştiinţifică, relaţia de tip cauză-efect este una mult mai complexă, pornind inclusiv de la faptul că variabila-cauză poate avea o altă semnificaţie decât aceea a unei condiţii necesare şi suficiente pentru a produce un anumit fenomen. Pentru a explica ideea cauzalităţii în ştiinţă, vom apela la exemplul oferit de Shadish, Cook şi Campbell (2002) cu privire la cauzele apariţiei unui incendiu într-o pădure. Un incendiu poate

avea cauze multiple, cum ar fi un chibrit aprins aruncat în iarbă, un foc de tabără nesupravegheat, un fulger etc. Niciuna dintre aceste posibile cauze nu este necesară (de exemplu, un incendiu poate apărea şi în situaţii în care nu este vorba despre un foc de tabără), nici suficientă (de exemplu, un foc de tabără nesupravegheat, pentru a se transforma în incendiu, ar trebui să fie însoţit de o serie de condiţii favorabile explicite, cum ar fi prezenţa unor lemne sau frunze uscate, sau implicite, cum ar fi prezenţa oxigenului pentru combustie). Astfel cauza potenţială (chibritul aprins aruncat în iarbă, focul de tabără nesupravegheat, fulgerul etc.) face parte dintr-o constelaţie de factori necesari pentru apariţia efectului – incendierea pădurii.

Din acest punct de vedere, aceste cauze potenţiale au fost denumite condiţii „inus” de către Mackie (1974 *apud* Shadish *et al.*, 2002), adică o condiţie insuficientă, neredundantă, ca parte a unei condiţii nenecesare, dar suficiente. Să ilustrăm aceste cuvinte abstracte prin exemplul incendierii pădurii. Lăsarea nesupravegheată a unui foc de tabără aprins reprezintă o condiţie insuficientă pentru a porni un incendiu, în absenţa altor factori favorizatori, dar este o acţiune neredundantă, contribuind într-o manieră unică la explicarea unei părţi însemnate dintre incendiile declanşate de-a lungul timpului, alături de alte cauze potenţiale posibile. Totuşi, lăsarea nesupravegheată a focului de tabără poate fi privită ca parte a unui întreg (alături de existenţa de lemne şi frunze uscate, absenţa ploii etc.) ce reprezintă o condiţie suficientă pentru pornirea incendiului, dar nu şi necesară, deoarece pot exista şi alţi factori responsabili pentru apariţia incendiului. O denumire alternativă pentru această perspectivă despre cauzalitate este oferită de Riegelman (1979), prin conceptul de condiţie contributivă. Din acest punct de vedere, o cauză poate fi contributivă dacă modificarea variabilei-cauză afectează variabila-efect,

cu precizarea că nu întotdeauna o modificare a variabilei-cauză se soldează cu modificări ale variabilei-efect, la fel cum nici orice caz protejat de acţiunea variabilei-cauză nu este scutit de apariţia efectului. De aceea, în psihologie majoritatea relaţiilor de tip cauză-efect nu descriu relaţii deterministe (în care variabila-cauză este necesară şi suficientă pentru a produce un anumit efect), ci se referă la relaţiile de tip cauză-efect ca la nişte condiţii contributive, pe care evaluările sistematice bazate pe metaanaliză (vezi capitolul 7) încearcă să le desluşească, subliniind variabilele de tip mediator sau moderator care descriu dinamica relaţiei dintre variabilele de tip cauză şi cele de tip efect.

De asemenea, o altă distincţie utilă pentru domeniul psihologiei este aceea de a diferenţia între relaţiile cauzale cu rol descriptiv şi relaţiile cauzale cu rol explicativ (Shadish *et al.*, 2002). Dintre cele două, mult mai uşor de demonstrat sunt relaţiile cauzale de tip descriptiv, cum ar fi evidenţierea faptului că muzica lui Mozart poate fi eficientă în îmbunătăţirea performanţelor la probe cognitive spaţiale. Pentru acest lucru este suficient să realizăm un studiu experimental care să compare nivelul performanţei la sarcini spaţiale al celor din grupul care a beneficiat două ore de muzica lui Mozart cu cel al indivizilor incluşi într-un grup placebo, care a beneficiat de muzica clasică a altor compozitori, dar prezentată ca fiind muzica lui Mozart. Cu alte cuvinte, este suficient să demonstrăm că variaţii ale variabilei independente (cauză) efectuate într-un context controlat, precum cel specific studiilor experimentale, produc modificări la nivelul efectului urmărit în direcţia aşteptată. În schimb, mult mai dificil de demonstrat este cazul unei relaţii cauzale de tip explicativ, în care se încearcă să se stabilească motivul pentru care relaţia cauzală este una funcţională. Astfel, dacă într-o relaţie cauzală descriptivă obiectivul fundamental este acela de a verifica dacă există o

relaţie de tip cauză-efect, într-o relaţie cauzală explicativă obiectivul fundamental devine acela de a identifica motivele care stau în spatele relaţiei cauzale (care anume dintre caracteristicile asociate muzicii lui Mozart constituie elementul activ, responsabil pentru îmbunătăţirea performanţei la sarcini spaţiale – o creştere a capacităţii de atenţie concentrată asupra sarcinii de rezolvat; o creştere a activităţii corticale pentru ariile cerebrale responsabile pentru procesarea informaţiilor spaţiale; inducerea unei dispoziţii afective pozitive care descarcă sistemul cognitiv de alocarea unor resurse spre direcţii nerelevante cum ar fi cazul anxietăţii de performanţă etc.). Detalii despre distincţia dintre studiile cu obiective cauzal-descriptive şi cele cu obiective cauzal-explicative[2] au fost oferite în capitolul anterior (secţiunea 3.3.5).

4.1.2. Controlul prin manipularea variabilei independente

Prin manipularea variabilei independente (VI), se înţelege decizia cercetătorului de a varia nivelul variabilei-cauză pentru a observa efectele apărute pornind de la premisele bazale conform cărora pentru a putea vorbi despre o relaţie de tip cauză-efect trebuie să observăm că: (a) modificările la nivelul VI se soldează cu modificări la nivelul variabilei dependente (VD); (b) modificările la nivelul VI le precedă pe cele apărute la nivelul VD. De pildă, un cercetător interesat să evidenţieze relaţia dintre consumul de calorii şi greutatea corporală poate decide să varieze consumul de calorii (de exemplu, grupul experimental să consume jumătate din numărul de calorii consumat de grupul de control). Un cercetător care doreşte să compare eficienţa a două tipuri de psihoterapie, cognitiv-comportamentală *vs* familial-sistemică, în combaterea depresiei

majore va decide să varieze tipul de psihoterapie urmată de către participanţii la studiu, astfel încât unii dintre aceştia să urmeze şedinţe de terapie cognitiv-comportamentală, iar alţii şedinţe de terapie familial-sistemică. Se observă, aşadar, intervenţia activă a cercetătorului, care poate decide care vor fi aspectele supuse manipulării. De exemplu, chiar în contextul în care participanţii sunt supuşi aceluiaşi tip de stimuli, manipularea poate avea în vedere instructajul oferit, timpul avut la dispoziţie pentru rezolvarea sarcinii, nivelul beneficiilor oferite participanţilor la studiu etc. Important din acest punct de vedere este să fie manipulat, pe rând, câte un singur element diferenţiator per grup experimental. Dacă se încalcă acest principiu, interpretarea rezultatelor are de suferit, deoarece ajungem să vorbim despre variabile confundate. De exemplu, dacă se compară eficienţa terapiei cognitiv-comportamentale (TCC) aplicată individual cu terapia psihodinamică (P) de grup pentru a ameliora anxietatea socială, nu vom putea interpreta acurat rezultatele cu privire la care dintre cele două forme de psihoterapie ar fi mai eficientă, deoarece este posibil ca diferenţele la nivel de efect sesizate să fie cauzate de celălalt factor distinctiv dintre grupe – modalitatea de implementare: individuală, respectiv în grup. Soluţia pentru asemenea situaţii, în care se doreşte manipularea a mai mult decât o singură dimensiune, presupune trecerea de la un plan de bază unifactorial la planuri multifactoriale. De exemplu, într-un plan bifactorial (conţine două variabile independente) se poate răspunde adecvat la mai multe întrebări precum: (a) „care formă de psihoterapie urmată (TCC *vs* P) este mai eficientă?”; (b) „care modalitate de implementare (individual *vs* de grup) este mai eficientă?”; (c) „există un efect de interacţiune între aceşti factori care să ne ajute la interpretarea nuanţată a efectului – reducerea nivelului de anxietate socială”.

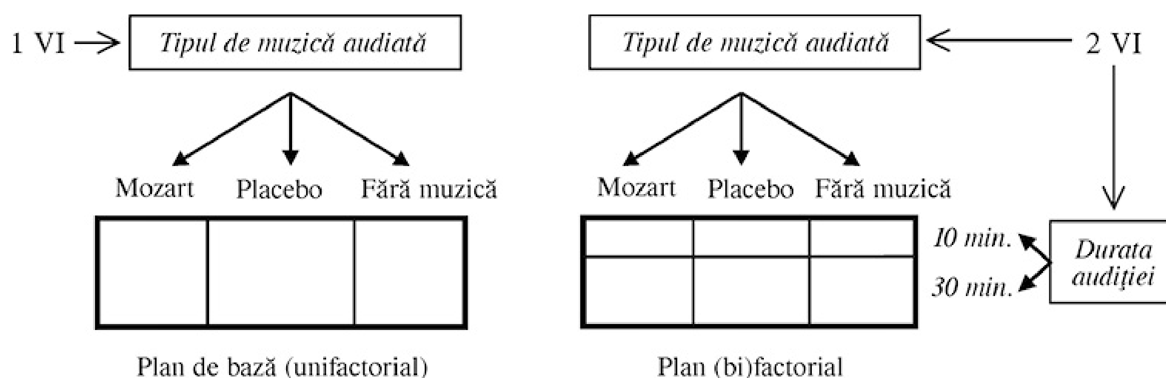


Figura 4.1. *Designuri experimentale de bază (stânga) şi factoriale (dreapta)*

Aceste designuri factoriale pot fi obţinute nu numai prin încrucişarea a două sau mai multe VI manipulate de către cercetător (de exemplu, tipul de muzică audiată şi durata audiţiei) (*completely randomized design)*, ci şi prin intersectarea unei variabile manipulate experimental (de exemplu, tipul de muzică ascultată) cu o variabilă de tip etichetă, ce descrie diferenţe existente în mod natural între participanţi (cum ar fi sexul participanţilor: bărbaţi *vs* femei) (*blocked randomized design)*.

În sfârşit, manipularea VI se poate face în două feluri majore: prin apelul la un design de tip intergrup (*between-subjects design*), respectiv prin apelul la un design de tip intragrup, de tipul măsurătorilor repetate (*within-subjects design*). În cazul designului de tip intergrup fiecare participant trece printr-o singură condiţie experimentală (unii ascultă muzică de Mozart, alţii aşteaptă în linişte etc.), astfel încât fiecare participant este testat o singură dată. Situaţia opusă este reprezentată de designul intragrup în care fiecare participant trece prin toate condiţiile experimentale, performanţele sale fiind măsurate în mod repetat. De exemplu, dacă ar fi să ne raportăm la designul de bază prezentat în partea stânga a figurii 4.1, aceeaşi participanţi vor fi testaţi după fiecare situaţie în care sunt supuşi: ascultând muzică de Mozart, ascultând muzica unui alt compozitor, dar prezentată ca fiind a lui Mozart, respectiv aşteptând în

linişte. Desigur, în cazul unor designuri de tip factorial (adică a prezenţei a cel puţin două VI), se pot întâlni şi situaţii mixte (design mixt), dacă una dintre variabilele independente este manipulată intergrup, iar cealaltă variabilă independentă este manipulată într-o manieră de tip intragrup.

4.1.2.1. Probleme uzuale legate de manipularea VI prin designuri de tip intragrup

Dintre tipurile de design amintite mult mai răspândite sunt cele de tip intergrup, în ciuda faptului că necesită un număr mult mai mare de participanţi decât designurile de tip intragrup. Acest lucru este cauzat de unele neajunsuri asociate manipulării variabilelor independente într-un design cu măsurători repetate, aspecte descrise în tabelul de mai jos.

Tabelul 4.1. *Riscuri privitoare la validitatea internă a rezultatelor în experimente cu un design intragrup (adaptare după Pelham şi Blanton, 2007)*

Denumire	*Descrierea problemei*	*Soluţii existente*
Reportarea efectului[1] (efectul de ordine)	Efectul apare atunci când răspunsul la stimulul anterior influenţează răspunsul la stimulul următor, fără a se dori explicit acest lucru. Un exemplu de efect de tip *carry-over* este cazul în care se solicită participanţilor să aprecieze cât de gravă este încălcarea regulii de a purta centură de siguranţă, după ce în prealabil persoanele au fost chestionate despre gravitatea comiterii unui omor. Participanţii incluşi în această situaţie vor aprecia că este mai puţin gravă abaterea de a nu purta centură de siguranţă prin contrast cu gravitatea omorului. În schimb, abaterea este considerată a fi mai gravă de către participanţii care ar răspunde	Atât pentru reducerea/evitarea efectului de reportare, cât şi pentru evitarea distribuirii într-o manieră inegală a consecinţelor oboselii sau implicării scăzute în sarcină există o soluţie reparatorie, denumită contrabalansare. Pentru un număr redus de probe de completat (2-3 probe) ideal ar fi să se apeleze la contrabalansarea totală, ceea ce înseamnă că ordinea în care se răspunde la probe este diferită de la o persoană la alta, acoperindu-se toate combinaţiile posibile. De exemplu, dacă există două probe/sarcini (A şi B), acestea

	mai întâi la întrebarea respectivă şi abia mai apoi la cea legată de infracţiunea de omor. Dincolo de efectele de reportare bazate pe efectul de contrast sau de cele bazate pe efectul de similaritate, se poate vorbi despre un efect de reportare în urma exersării (practicării repetate a sarcinii), astfel încât sarcinile aflate la final să fie realizate mai rapid (facil), dacă se aseamănă cu primele.	pot fi prezentate randomizat, fie în ordinea AB, fie în ordinea BA. În cazul a trei probe (ABC) combinaţiile posibile sunt ABC, ACB, BCA, BAC, CAB şi CBA. Dacă numărul de probe este unul mai ridicat, devine imposibil să apelăm la contrabalansarea totală deoarece numărul de combinaţii posibile depăşeşte adesea numărul de participanţi într-un studiu (de exemplu, dacă ar fi să combinăm 5 probe într-o manieră random unică vom avea 5! combinaţii posibile, adică 120 de combinaţii unice). În aceste cazuri se apelează la soluţii de contrabalansare parţială, dintre care cea mai cunoscută este pătratul latin (detalii în Pelham şi Blanton, 2007). În situaţia combinării a patru probe: A, B, C şi D este necesar ca acestea să fie aplicate randomizat în patru succesiuni diferite: ABCD, BADC, CDAB şi DCBA.
Oboseala/lipsa de motivare	Dacă sarcinile pe care le are de rezolvat o persoană într-un design intragrup solicită un efort îndelungat din partea respondenţilor, este posibil să se instaleze oboseala, fapt ce ar afecta performanţa în sarcinile aflate spre finalul setului de probe de completat. De asemenea, în cazul persoanelor cu motivaţie scăzută se poate instala dezinteresul şi poate creşte riscul de completare la întâmplare a probelor aflate spre finalul setului de sarcini primite.	
Creşterea riscului ca participanţii să intuiască ipotezele studiului	Ideal ar fi ca participanţii la un studiu să nu intuiască ce se urmăreşte exact prin studiul respectiv pentru a nu fi afectaţi în răspunsurile lor de aceste presupoziţii (în limba engleză acest caz este cunoscut sub denumirea de *demand characteristics*). De exemplu, dacă participanţii efectuează exerciţii de relaxare în condiţii de linişte, apoi fac aceste exerciţii în condiţii în care ascultă muzică clasică, pot intui mai uşor că studiul vizează efectul pe care ambientul sonor îl are asupra eficienţei utilizării unor	Desigur, cea mai bună soluţie în acest context ar fi renunţarea la designul intragrup, în favoarea unuia integrup, deoarece este mai puţin probabil ca participanţii să intuiască despre ce era vorba în experiment, în contextul în care ar fi fost supuşi doar condiţiei de a exersa tehnicile de relaxare în linişte, spre exemplu. A doua modalitate de lucru valabilă atât pentru designuri de tip intergrup, cât şi pentru cele de tip intragrup

	tehnici de relaxare şi pot decide să răspundă într-un mod care să-l mulţumească pe experimentator (considerând că nivelul mai profund de relaxare este atins în contextul în care se ascultă muzică clasică).	este introducerea unor întrebări directe în partea de debriefing (finalul studiului) în care se solicită participanţilor să intuiască ipoteza studiului/aşteptările cercetătorului, însoţită de decizia de a-i elimina pe aceia care intuiesc ipotezele studiului din cauza riscului crescut ca răspunsurile acestora să fie distorsionate în favoarea conformării ipotezei.

În ciuda acestor neajunsuri ale unui design de tip intragrup[3], acesta are avantajul de a controla toate variabilele străine care ţin de individul examinat, fiind vorba de aceeaşi persoană supusă unor contexte experimentale diferite. Acest fapt conduce la o putere statistică superioară de a testa efectele urmărite, în condiţiile unui număr mai redus de participanţi. Spre deosebire de acest caz, în designurile experimentale de tip intergrup principala problemă care necesită rezolvare este tocmai controlul variabilelor străine ce ţin de caracteristicile participantului, în lipsa acestui control putând fi afectate concluziile cu privire la relaţia de tip cauză-efect. Pentru a rezolva această problemă cea mai potrivită soluţie este decizia de a distribui participanţii în grupele experimentale într-o manieră aleatorie (randomizată) pentru a egaliza (controla) astfel potenţialul efect distorsionant pe care variabilele confundate l-ar putea juca asupra relaţiei cauzale studiate. Înainte de a explicita în detaliu ce presupune distribuirea randomizată a participanţilor în grupele experimentale, se cuvine să precizăm o ultimă problemă ce poate afecta interpretarea rezultatelor obţinute în urma manipulării experimentale a VI, şi anume problematica variabilelor confundate.

4.1.2.2. Problema variabilelor confundate în cazul manipulării VI

Variabilele confundate (*confounding variables*, *nuisance variables* sau *hidden variables*) mai sunt cunoscute şi sub titulatura de variabile ascunse sau variabile străine, fiind vorba de acele variabile pe care cercetătorul nu le controlează într-un mod explicit, dar care pot afecta validitatea internă a unui studiu. De aceea se impun măsuri de control experimental sau statistic al efectului acestora. Detalii suplimentare despre acest tip de variabile sunt oferite în caseta 5.4 din capitolul următor.

Cele mai clare cazuri de variabile confundate pot fi evidenţiate în cazul studiilor de tip non-experimental. De exemplu, Pelham şi Blanton (2007) evidenţiază complexitatea fiinţei umane, fapt ce favorizează existenţa unor variabile confundate. De pildă, în urma unui studiu longitudinal ar putea reieşi următorul rezultat statistic corect: există o mortalitate mai crescută la indivizii care s-au tatuat decât în rândul celor care nu au niciun tatuaj. Mai mult, riscul de a deceda, în intervalul de zece ani avut sub observaţie, este cu atât mai ridicat cu cât indivizii posedă mai multe tatuaje. Desigur, o asemenea concluzie ar fi cauzată de existenţa unor variabile străine necontrolate (variabile confundate), pentru că este puţin probabil să poată fi controlate toate variabilele posibile, responsabile pentru un asemenea rezultat. De exemplu, o serie de factori ce cresc mortalitatea sunt asociaţi mai frecvent persoanelor care se tatuează. Astfel, persoanele care au mai multe tatuaje tind să adopte un stil de viaţă bazat pe: (a) un consum mai ridicat de tutun şi alcool; (b) o dietă cu multe grăsimi şi consum redus de fibre; (c) complianţă scăzută la tratamentele necesare în cazul unor condiţii cronice precum hipertensiune arterială sau diabet; (e) un stil mai agresiv (mai puţin precaut) de

conducere a autovehiculului; (f) o preferinţă pentru motociclete, ca mijloc de locomoţie; (g) nesocotirea unor norme şi reguli sociale etc. Desigur, dacă am fi controlat toţi aceşti factori posibili, asocierea dintre tatuare şi vârsta decesului ar fi devenit, probabil, nesemnificativă statistic, însă este greu de crezut că este fezabil să controlăm toate variabilele confundate posibile. Pe scurt, variabilele confundate sunt cele la care cercetătorul, fie nu s-a gândit, fie nu le-a putut controla, dar care prin natura lor (variază sistematic cu variabilele supuse asocierii – în cazul studiilor non-experimentale, respectiv cu VI şi VD – în cazul studiilor experimentale). Ele pot reprezenta sursa reală pentru efectul (relaţia) observat, afectând astfel validitatea unui studiu.

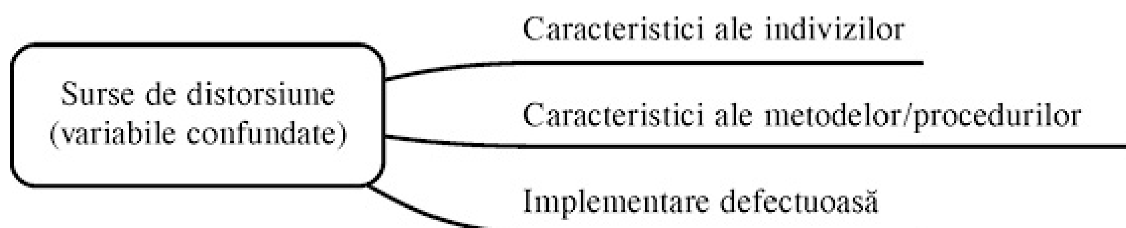


Figura 4.2. *Surse ale distorsiunilor determinate de variabile confundate*

4.1.2.2.1. Variabile confundate având drept sursă caracteristicile participanţilor

În studiile experimentale, aceste situaţii de variabile confundate legate de caracteristicile participanţilor sunt controlabile. În designurile de tip intergrup se apelează la distribuirea randomizată a subiecţilor în grupele experimentale – procedeu menit să conducă la constituirea unor grupe echivalente (similare) din perspectiva variabilelor potenţial confundate –, această metodă nu este una infailibilă, situaţie în care discutăm despre un bias (distorsiune) de selecţie. În cazul designului intragrup, variabilele confundate ce ţin de caracteristicile participanţilor sunt controlate integral

prin aplicarea procedurilor pe aceiaşi subiecţi. Secţiunea 4.1.3 prezintă pe larg procedura cea mai des întâlnită, aceea a alocării randomizate a participanţilor în grupele experimentale.

4.1.2.2.2. Variabile confundate având drept sursă metodele sau procedurile utilizate

Nu doar participanţii pot ascunde variabile confundate în sine, ci şi procedurile experimentale utilizate în manipularea VI (diferite modalităţi ale presupusei variabile-cauză). Să presupunem că un individ este interesat să urmărească dacă persoanele anxioase reţin mai bine cuvinte cu valenţă negativă (de exemplu, *bolnăvicios*), decât cuvinte cu valenţă pozitivă (de exemplu, *echitabil*). Din nefericire, aceste cuvinte nu diferă numai prin valenţa lor afectivă, ci şi prin nivelul de familiaritate (mai utilizat fiind *bolnăvicios*), prin nivelul de activare (*arousal*) – mai „viu” fiind tot cuvântul *bolnăvicios* etc. Dacă nu sunt controlate aspecte, precum gradul de familiaritate, lungimea cuvintelor, nivelul lor de expresivitate (intensitate), ar fi dificil să afirmăm că diferenţele observate în performanţa mnezică ar fi determinate de valenţa afectivă a cuvintelor, deoarece există şi alte dimensiuni prin care acestea diferă. Acelaşi lucru poate fi avut în vedere inclusiv în situaţii experimentale complexe. De pildă, atunci când se compară două forme de psihoterapie din perspectiva nivelului lor de eficienţă este important ca ambele forme să fie implementate de către un personal calificat, ca ambele să aibă niveluri procentuale similare de participanţi care renunţă la psihoterapie înainte de termenul convenit etc.

Pentru eliminarea efectului unor variabile străine (variabile pe care cercetătorul le-a luat în considerare ca potenţiali factori ce afectează relaţia cauză-efect) soluţiile de corectare sunt controlul statistic

(măsurarea variabilelor respective şi eliminarea efectului acestora prin tehnici statistice de tipul ANCOVA, corelaţii parţiale sau regresii ierarhice), respectiv controlul metodologic (de exemplu, dacă genul persoanei ar constitui o variabilă confundată se pot efectua studii doar pe femei sau pot fi formate grupele experimentale în proporţii egale pe sexe etc.). De asemenea, pentru eliminarea efectului unor variabile străine neluate în calcul sau necontrolate iniţial şi pentru decelarea impactului acestora la nivelul variabilei dependente, corectarea se poate face şi prin replicarea rezultatelor obţinute în contexte experimentale similare, dar nu identice, precum este exemplul descris în caseta 4.2.

Caseta 4.2. *Descrierea unui caz de variabilă confundată întâlnit în studiile experimentale derulate în scop explicativ (adaptare după Pelham şi Blanton, 2007)*

Dispoziţiile afective pozitive ne determină să fim mai altruişti?

Pentru a răspunde la această întrebare ce vizează identificarea unei relaţii de tip cauză-efect între dispoziţia afectivă şi comportamentul altruist, primele cercetări experimentale realizate într-un mod riguros au distribuit randomizat participanţii, fie în grupa celor care primeau un feedback pozitiv pentru activitatea depusă pentru a intensifica dispoziţia afectivă pozitivă a acestor persoane, fie în grupa celor care nu primeau un asemenea feedback. Datele obţinute din astfel de studii susţin ipoteza avută: cei care au primit un feedback pozitiv au relatat o dispoziţie afectivă pozitivă mai intensă decât grupul de control şi, ce este mai important, au fost mai altruişti în comportamentul lor (de exemplu, au acordat mai mulţi bani unor complici care „cerşeau” la ieşirea din locaţia experimentală).

Problema cu această manipulare experimentală este faptul că a acorda un feedback pozitiv unui interlocutor are ca efect nu numai inducerea unei dispoziţii afective pozitive, ci şi modificarea credinţei acestuia în competenţa sa de a face faţă provocărilor. Or, persoanele care se simt mai competente într-un domeniu sunt, de asemenea, mai predispuse să-i ajute pe cei care le solicită ajutorul, fiind astfel mai altruiste (Isen şi Levin, 1972 *apud* Pelham şi Blanton, 2007). De aceea, nu este clar dacă altruismul mai ridicat este o consecinţă a unei dispoziţii afective pozitive sau unei augmentări a încrederii celor chestionaţi în competenţa lor de a se descurca în sarcini.

Pentru a rezolva această dilemă, alţi cercetători au propus o manipulare diferită a dispoziţiei afective care să nu conducă şi la modificarea credinţei participanţilor în competenţa lor. În acest sens, celor din grupul experimental li s-au oferit prăjiturele în

timp ce rezolvau sarcina, pentru a observa dacă acest surplus indus de buna-dispoziţie conduce la augmentarea comportamentului altruist. Datele obţinute sprijină ipoteza că dispoziţia afectivă pozitivă (mai ridicată în grupul celor care au primit prăjituri) a condus la niveluri mai ridicate ale comportamentului altruist. Chiar dacă acest context experimental nu mai este unul care să conducă la modificări în competenţa autopercepută a respondenţilor, este posibil ca rezultatul obţinut să fie distorsionat de un alt aspect. Astfel, cei din grupul experimental au fost expuşi unui model de comportament prosocial (primind prăjituri din partea unor străini), în condiţiile în care expunerea la comportamente prosociale este o altă cauză a apariţiei unei conduite de tip altruist.

În sfârşit, pentru a rezolva şi această posibilă situaţie de variabile confundate, o altă manipulare propusă pentru inducerea unei dispoziţii afective pozitive a fost aceea de a manipula contextul experimental astfel încât cei din condiţia experimentală să găsească o monedă pierdută (nerevendicată) spre deosebire de cei din grupul de control care nu au avut un asemenea „noroc”. Şi în acest caz, rezultatele au indicat că persoanele din grupa cu o dispoziţie afectivă pozitivă mai ridicată (norocoşii) au manifestat un comportament prosocial mai intens decât cei din grupa de control. Acest design are însă meritul de a nu conduce nici la modificări în competenţa autopercepută a candidaţilor, nici la situaţii de expunere la un model prosocial. Prin urmare, atât prin replicarea rezultatelor în contexte experimentale diferite, cât mai ales prin identificarea unui design de cercetare care să elimine posibile explicaţii alternative (variabile confundate), există un sprijin consistent pentru a afirma că o dispoziţie afectivă pozitivă creşte comportamentul de tip altruist.

Desigur, pe lângă soluţia principală de a manipula experimental VI de interes astfel încât aceasta să înceteze a mai fi afectată de o potenţială variabilă confundată, există întotdeauna posibilitatea replicării unui studiu, cu includerea unei măsurători pentru variabila confundată analizată. De exemplu, în contextul repetării unui studiu experimental bazat pe inducerea unei dispoziţii afective pozitive ca urmare a unui feedback pozitiv, se poate include suplimentar drept covariabilă măsurarea competenţei autopercepute prin intermediul unei scale validate din punct de vedere psihometric, astfel încât analizele statistice efectuate să evidenţieze eventualele diferenţe dintre grupul experimental şi cel de

control, în contextul în care este controlat statistic nivelul competenţei autopercepute (în situaţia ipotetică în care toţi respondenţii ar avea acelaşi nivel al competenţei autopercepute).

4.1.2.2.3. Variabile confundate având drept sursă implementarea distorsionată a studiului

În această categorie sunt incluse toate acele cauze potenţiale care vizează direct modul în care este implementat un anumit studiu. De exemplu, în cazul unui experiment pot apărea distorsiuni cauzate de: (a) lipsa de standardizare a instructajului oferit participanţilor la studiu (variaţii în lungimea şi nivelul conversaţiilor cu participanţii premergătoare studiului, instructaj scurtat în cazul unor participanţi care par a prinde repede ce au de făcut etc.); (b) distorsiuni cauzate de caracteristicile şi aşteptările cercetătorului/asistentului de cercetare (caracteristici demografice şi stil de lucru diferit al asistenţilor de cercetare; comportament diferenţiat faţă de participanţi în funcţie de condiţia lor experimentală – cum ar fi niveluri diferite de entuziasm şi de atenţie acordată participanţilor); (c) distorsiuni legate de caracteristicile sarcinii experimentale (*demand characteristics*) (dacă sarcinile sunt prea transparente, participanţii vor intui ipotezele cercetării şi se vor comporta predominant într-un mod care să mulţumească experimentatorul/confirme aşteptările acestuia privind cel mai potrivit mod de a răspunde sau a se comporta).

Pentru a elimina asemenea posibile distorsiuni[4], pot fi implementate o serie de măsuri precum: (a) standardizarea procedurilor de testare (instructaj oferit de către calculator, utilizarea unor „falşi subiecţi” pentru a verifica gradul de respectare a instrucţiunilor de examinare); (b) utilizarea unui design de cercetare de tip orb (în care participanţii să nu cunoască exact condiţia în care au fost distribuiţi – grupuri placebo – sau

în care cei care efectuează evaluările nu cunosc cărui grup aparţin participanţii testaţi) sau dublu orb (în care atât participanţii, cât şi asistenţii de cercetare, cel puţin cei care efectuează măsurătorile de interes, nu cunosc cărui grup experimental aparţin indivizii testaţi); (c) utilizarea unui interviu sau a unui chestionar la finalul cercetării (*debriefing*) pentru a detecta cazurile potenţiale de a fi afectate de caracteristicile sarcinii (indivizii care au intuit corect ipoteza vizată de cercetare şi relaţia dintre sarcinile primite spre rezolvare şi ipoteza studiului).

4.1.2.3. Problema artefactelor în cazul manipulării VI

Spre deosebire de variabilele confundate care variază atât cu VI, cât şi cu VD, artefactele reprezintă o serie de distorsiuni constante în cadrul studiului experimental. Prin efectul lor constant, ele nu afectează validitatea internă a unui studiu, ci pe cea externă, deoarece rezultatul obţinut între VI şi VD este valabil numai în contextul în care artefactul respectiv este prezent. În categoria artefactelor intră de obicei acele variabile pe care cercetătorul nu este interesat să le studieze, dar care prin efectul lor afectează validitatea externă a concluziilor unui studiu.

Un exemplu de artefact este reactivitatea participanţilor într-un studiu experimental de laborator. Aceştia pot adopta, pentru că sunt observaţi un anumit comportament care să nu fie replicabil în contexte sociale ecologice. Faptul că majoritatea studiilor experimentale se derulează în condiţii de laborator pentru a controla diferite variabile străine, nu face altceva decât să replice efectul artefactului în contexte în care participantul realizează faptul că este monitorizat comportamentul său.

Creşterea realismului situaţiei experimentale poate contribui la diminuarea efectului pe care artefactul reactivităţii participanţilor îl joacă în dinamica dintre VI şi VD. Un alt exemplu de artefact este mediul cultural (de exemplu, unele efecte pot fi replicate doar în societăţi individualiste, orientate spre competiţie, dar nu şi în medii culturale estice, cum este asocierea dintre percepţia unicităţii – „sunt o persoană specială" şi stima de sine). Dacă în societăţile individualiste există o asociere pozitivă între percepţia unicităţii şi stima de sine, în cele colectiviste această relaţie este una negativă.

Iată un alt exemplu: un consilier creştin este interesat să studieze rolul benefic al religiei şi orientării către spiritualitate la bolnavii de cancer; acest rol pozitiv a fost replicat şi în alte studii efectuate pe credincioşi. Statutul de persoană credincioasă poate reprezenta în aceste cazuri un artefact. Pentru a controla influenţa acestui factor, ar fi interesant să vedem dacă religia şi practicile spirituale sunt mecanisme benefice de coping la bolnavii de cancer care se declară atei. În sfârşit, artefacte pot rezulta şi din favorizarea (standardizarea) unor anumite proceduri de lucru în studiile experimentale. De exemplu, în cazul condiţionării afective (transmiterea valenţei afective de la un stimul necondiţionat spre un stimul condiţionat, cu valenţă iniţială neutră), se pare că unul dintre cele mai importante mecanisme pentru apariţia efectului este conştientizarea asocierii dintre stimulul necondiţionat şi cel condiţionat. Însă această concluzie are o valabilitate limitată de vreme ce peste 90% dintre studiile din această zonă utilizează doar măsuri explicite pentru a aprecia preferinţa pentru stimulul condiţionat (Hoffman *et al.*, 2011). Dacă s-ar lua în calcul măsuri implicite ale preferinţei pentru stimulul condiţionat mecanismul pentru apariţia schimbării ar putea fi unul diferit de cel al conştientizării asocierii dintre stimuli.

Spre deosebire de cazul variabilelor confundate, artefactele sunt mult mai dificil de identificat şi de corectat. Pentru a evita situaţia în care relaţia dintre VI şi VD există doar în contextul existenţei unor artefacte, ideal ar fi să se replice rezultatele prin varierea situaţiilor experimentale pe diferite dimensiuni relevante, pentru a creşte validitatea externă a concluziilor şi, îndeosebi, validitatea ecologică a rezultatelor obţinute.

Caseta 4.3. *Studiu de caz*

Pornind de la datele existente în literatura de specialitate, terapia prin muzică poate fi o metodă eficientă de reducere a anxietăţii şi, în general, de reducere a stresului, atât la adulţi, cât şi la copii (Bradt şi Dileo, 2009; Klassen, Liang, Tjosvold, Klassen şi Hartling, 2008 etc.). Tipul de muzică pare a fi un factor important, cele mai bune rezultate din studiile făcute pe populaţie americană fiind obţinute pe muzică clasică, muzică din genul New Age şi latino-jazz, iar cele mai puţin eficiente fiind obţinute pe muzică rock (*apud* Pfister, Jones şi Cheek, 2011)

O cercetare efectuată în România pe această temă (Gherman, 2008) a urmărit să replice efectul pozitiv al muzicii asupra reducerii anxietăţii pe un lot de pacienţi cardiaci supuşi unei intervenţii invazive (stentare carotidiană). Au fost avute în vedere trei tipuri de muzică, audiate cu acordul participanţilor la căşti, în timpul derulării procedurii de stentare: fie muzică clasică, fie muzică latino-jazz, fie folclor (muzică populară). Rezultatele studiului experimental au arătat că nivelul cel mai scăzut de anxietate a fost obţinut în condiţia de ascultare a muzicii populare, în timp ce s-a înregistrat chiar o creştere a nivelului de anxietate postoperatorii faţă de pretest, în grupul celor care au ascultat muzică latino-jazz, respectiv clasică.

Ce cauze (explicaţii) sunt posibile pentru divergenţa rezultatelor, pornind de la premisa că toate studiile au avut un nivel calitativ superior (randomizare, fără alte artefacte induse de experimentator)?

1. ***Probleme privind validitatea de construct a studiului***

Divergenţa este cauzată de prezenţa unor variabile confundate în modul de manipulare a VI: *familiaritatea cu genul muzical; preferinţa pentru un anumit gen muzical.*

Este posibil ca responsabilitatea pentru scăderea nivelului de anxietate să nu aparţină caracteristicilor melodice (tonalitate, ritm), ci gradului de familiarizare a persoanei cu genul muzical respectiv şi/sau preferinţele sale muzicale. Intervenţia la care sunt supuşi participanţii este una invazivă, prin urmare inductoare a unui cadru anxiogen. Apelul la

stimuli familiari sau plăcuţi pare a fi chiar o modalitate rezonabilă de coping în asemenea situaţii.
Dacă această explicaţie este adevărată, atunci este afectată validitatea internă a studiilor anterioare, iar concluzia că muzica clasică sau cea latino-jazz sunt cele mai eficiente în reducerea anxietăţii şi stresului ar fi hazardată. În acest context ar fi nevoie de studii suplimentare cu scopul de a stabili relaţia dintre familiaritate, preferinţă sau tipul de muzică ca elemente active pentru reducerea anxietăţii/stresului.

2. ***Probleme privind validitatea externă a studiului***
Divergenţa este cauzată de prezenţa a două posibile artefacte în studiile anterioare: (a) mediul cultural relativ omogen (peste 90% din studiile anterioare sunt realizate în SUA, pe loturi preponderent formate din studenţi); (b) momentul introducerii calupului muzical.

Astfel, eşantionul din România diferă semnificativ faţă de cel utilizat în studiile anterioare, atât din perspectiva gusturilor muzicale, cât şi din perspectiva nivelului de studii, deoarece participanţii incluşi în studiul din România aparţin într-o proporţie de peste 75% celor care au cel mult 8 clase de şcoală, un grup-ţintă care, foarte probabil, este familiarizat şi îndrăgeşte muzica populară, dar în acelaşi timp manifestă reticenţă faţă de stilul muzicii clasice sau latino-jazz.
În privinţa momentului introducerii muzicii, există alte diferenţe între studiul realizat în România şi majoritatea celor realizate anterior. În studiile anterioare, participanţii ascultau muzică într-un context preoperator (sala de pregătire pentru intervenţie), în timp ce în studiul efectuat în România participanţii ascultau muzica la căşti în timpul intervenţiei de stentare carotidiană).
Dacă aceste explicaţii sunt adevărate, atunci validitatea internă a studiilor rămâne una corectă (muzica clasică şi cea latino-jazz pot fi eficiente în reducerea anxietăţii), dar dobândesc o valabilitate limitată (de exemplu, în mediul cultural american/în cazul persoanelor cu studii superioare etc.), fiind în acest fel afectată validitatea externă, adică gradul de generalizare a rezultatelor.

În sfârşit, divergenţa dintre rezultate poate avea drept sursă combinaţia elementelor menţionate anterior, situaţie în care ar fi afectată, atât validitatea de construct, cât şi cea externă. Desigur, fiind vorba de un singur studiu cu rezultate contrare, ar fi nevoie de replicarea acestor rezultate, preferabil de către echipe de cercetători diferiţi, pentru ca dovezile de combatere a concluziilor iniţiale să fie considerate unele solide. Dacă asemenea studii ar replica rezultatele obţinute în studiul din România, s-ar ajunge la revizuirea asumpţiei că numai anumite tipuri de muzică sunt eficiente în combaterea anxietăţii.

4.1.3. Controlul variabilelor străine prin randomizare

Atunci când se compară performanţele grupului experimental cu cele ale grupului de control, este important ca acest lucru să se realizeze în condiţiile în care riscul existenţei unei distorsiuni (bias) de selecţie a participanţilor să fie minim. De asemenea, este important ca riscul apariţiei unor modificări temporale în mod diferenţiat între grupe să fie minim. Avem nevoie de acest lucru pentru a ne putea încrede că eventualele diferenţe observate sunt determinate de manipularea variabilei independente (cauza acceptată a diferenţelor existente între grupe) şi nu de existenţa unor diferenţe observate în cazul altor variabile străine (diferenţe existente între grupul experimental şi cel de control).

O modalitate de a exercita controlul asupra variabilelor străine, cunoscută sub denumirea de control statistic al variabilelor, se face prin intermediul unor tehnici statistice cum ar fi regresia multiliniară ierarhică sau analiza de covarianţă (ANCOVA). Cu ajutorul acestor tehnici se poate elimina efectul variabilelor străine luate în calcul asupra relaţiei cauzale de interes. Problema cu acest tip de control al variabilelor străine este aceea că nu pot fi avute în vedere toate variabilele posibile care pot afecta relaţia de tip cauză-efect studiată. De exemplu, într-un studiu care compară eficienţa unei anumite metode alternative de predare a matematicii comparativ cu metoda uzuală de predare din perspectiva rezultatelor obţinute la bacalaureat la disciplina matematică pot exista o multitudine de diferenţe posibile între elevi cum ar fi: (a) niveluri diferite de cunoştinţe despre matematică între grupul experimental şi cel de control; (b) niveluri diferite ale abilităţilor cognitive generale (de inteligenţă) ce afectează ritmul învăţării şi retenţia cunoştinţelor noi; (c)

niveluri diferite de motivaţie pentru performanţă şcolară; (d) niveluri diferite de expertiză din partea profesorilor care predau cele două metode comparate; (e) niveluri diferite ale climatului şcolar; (f) politici diferite cu privire la absenteismul şcolar; (g) niveluri diferite de statut socioeconomic ale părinţilor.

Deşi pot fi controlate statistic multe asemenea variabile importante pentru a elimina efectul acestora asupra relaţiei dintre variabila independentă (metoda de predare) şi cea dependentă (performanţa la bacalaureat), este improbabil că vor putea fi controlate statistic toate variabilele care au impact asupra relaţiei dintre VI şi VD. Soluţia optimă pentru a controla efectul potenţial al tuturor variabilelor relevante asupra relaţiei dintre VI şi VD nu este aceea a controlului statistic (măsurarea tuturor acestor variabile), ci a controlului exercitat prin design, adică prin alocarea randomizată a subiecţilor în grupele experimentale.

Acest lucru presupune faptul că indivizii incluşi în studiu nu sunt comparaţi în cadrul grupurilor naturale din care fac parte (de exemplu, copii de la grădiniţe de stat *vs* copii de la grădiniţe particulare) şi nici măcar nu au posibilitatea de a-şi alege ei grupul din care să facă parte. În schimb, indivizii incluşi în studii experimentale consimt informat să lase în responsabilitatea cercetătorului sarcina de a-i distribui aleatoriu (ca într-o loterie) într-unul din grupele experimentale, chiar fără a li se comunica decât la finalul studiului natura reală a grupului din care au făcut parte. Principala problemă pe care o rezolvă alocarea randomizată a participanţilor în grupele experimentale este eliminarea distorsiunii de selecţie a participanţilor, cu condiţia ca grupurile formate să fie suficient de mari (de exemplu, peste 40 de persoane în fiecare grup experimental).

Pentru a înţelege importanţa acestui aspect, vom apela la un exemplu din ştiinţele medicale. În mod obişnuit se recunoaşte impactul pozitiv al

consumului de suplimente alimentare precum vitamina C sau vitamina E ca factori protectivi pentru imunitate etc. Aceste concluzii se bazează însă pe studiul unor grupuri natural constituite (fără alocare randomizată), bazată pe compararea celor care consumă asemenea suplimente alimentare în mod voluntar cu cei care nu apelează la aceste suplimente. Asemenea rezultate eşuează însă în a ţine seama de alte diferenţe existente între cele două grupe comparate (de exemplu, cei care folosesc asemenea aditivi tind să apeleze de asemenea la o dietă alimentară mai corectă, să fie mai puţin predispuşi la consumul de alcool şi tutun etc.). Prin alocarea randomizată a participanţilor într-un asemenea studiu se apelează la o populaţie selecţionată (de exemplu, persoane care nu consumă alcool decât ocazional, care nu fumează, nu au un istoric al unor boli cronice, nu consumă aditivi alimentari) pe care apoi îi împarte aleatoriu în două grupe: unii care vor consuma regulat vitamina C şi cei care cred că vor consuma regulat vitamina C (primind pastile placebo). Dacă în aceste condiţii stricte de control există diferenţe între cei care consumă vitamina C şi cei care nu consumă vitamina C, am putea spune că vitamina C are un efect benefic asupra sănătăţii din perspectiva reducerii riscului morbidităţii. Surprinzător, poate, rezultatele obţinute în asemenea condiţii de control nu confirmă credinţele populare. Astfel, o metaanaliză de studii experimentale realizată de Bjelakovic *et al.* (2007) a arătat că cei care consumă antioxidanţi precum vitaminele A, E sau betacaroten prezintă un risc crescut de mortalitate comparativ cu cei din grupurile de control – tip placebo), iar în cazul celor care consumă suplimente precum selenium sau vitamina C nu există diferenţe semnificative statistic cu privire la riscul mortalităţii între grupul experimental şi cel de control. Aşadar, comparaţiile între grupul experimental şi cel de control au sens în a fi interpretate corect doar în

absenţa unei distorsiuni în selecţia participanţilor, iar acest lucru poate fi realizat, fie prin apelul la un design intragrup (acolo unde este posibil), fie prin alocarea randomizată a participanţilor în grupele experimentale, în cazul unor designuri de tip intergrup.

Merită menţionată aici distincţia conceptuală importantă dintre selecţia aleatorie a unui eşantion de persoane din populaţia de interes (*random selection*) şi alocarea randomizată a subiecţilor în grupe de studiu (*random assignment*). Primul caz are în vedere ceea ce se întâmplă adesea în sondajele de opinie (chestionarea unui număr de 100 de persoane dintr-o comună cu un total de 5.200 de locuitori prin selectarea aleatorie a persoanelor chestionate – aspect ce potenţează validitatea externă a unui studiu). Cel de-al doilea caz nu-şi propune să selecteze aleatoriu eşantionul chestionat, acesta putând fi unul de convenienţă (cum ar fi 100 dintre locuitorii satului prezenţi la o întâlnire organizată de primărie), ci îşi propune să îi distribuie aleatoriu pe participanţi în grupele experimentale pentru a-i chestiona în două situaţii diferite: unii fiind rugaţi să aprecieze activitatea primăriei, în regim de anonimat, iar alţii, în contextul în care li se solicită să-şi dezvăluie identitatea înainte de a-şi exprima opiniile. Alocarea randomizată a participanţilor în grupele de studiu este o condiţie esenţială în designurile intergrup pentru a asigura un nivel ridicat al validităţii interne a cercetării.

4.1.3.1. Sugestii practice pentru alocarea randomizată a participanţilor

A aloca randomizat participanţii în grupele experimentale înseamnă a acorda şanse egale tuturor participanţilor de a fi incluşi în oricare dintre condiţiile experimentale ale studiului. Într-o variantă minimală, acest

lucru presupune apelul la unele metode clasice precum randomizarea sistematică (de exemplu, într-un design cu trei grupe se utilizează un pas de trei, astfel încât indivizii cu numărul de ordine [înscriere] 1, 4, 7, 10 etc. sunt distribuiţi în primul grup, cei cu numerele 2, 5, 8, 11 etc. sunt distribuiţi în cel de-al doilea grup, iar indivizii cu numerele 3, 6, 9, 12 etc. sunt incluşi în cel de-al treilea grup experimental). Alte variante rudimentare de alocare randomizată a indivizilor presupun utilizarea unor criterii neutre de selecţie (cei născuţi în luni pare *vs* impare; cei care au iniţiala numelui de familie de la A la I *vs* de la J la Z; ultima cifră pară sau impară din CNP-ul participantului etc.).

Într-o lume ideală, asemenea procedee ar fi considerate adecvate pentru o alocare randomizată de calitate. Din nefericire, însă, aceste criterii pot conduce la conduite conştiente sau inconştiente soldate cu o distorsiune de selecţie. De exemplu, în mod conştient, unui participant îi poate fi rezervat un număr care să-l distribuie într-o anumită grupă (de exemplu, dacă la interviul iniţial participantul pare a fi un individ recalcitrant şi rezistent la sugestii, poate ajunge să fie inclus în grupul de control, deoarece se anticipă că nu vor fi sesizate modificări mari între test şi retest). De asemenea, în cazul în care se cunoaşte deja grupul în care va fi distribuit un anumit participant (pentru că iniţiala numelui său de familie începe cu litera V) acest lucru poate influenţa evaluarea iniţială a îndeplinirii/neîndeplinirii criteriilor de eligibilitate de a fi inclus în studiu, mai ales în contextul în care criteriile de eligibilitate sunt vagi. De pildă, individul poate fi considerat drept eligibil dacă urmează a fi distribuit în grupul în care se estimează că ar avea şanse să „ajute” la susţinerea ipotezei de lucru, respectiv să rămână neselectat (declarat neeligibil) dacă se anticipă faptul că nu este potrivit intervenţiei alocate (Torgerson şi Torgerson, 2008).

De aceea, o variantă optimizată a procesului de alocare randomizată a participanţilor în grupele experimentale are în vedere ascunderea procesului de randomizare, astfel încât aceasta să aibă loc *fără a se cunoaşte* a priori *în ce grupă urmează să fie inclusă o anumită persoană testată*. Tabelul 4.2 sintetizează principalele recomandări aplicative privind practica alocării randomizate în grupe experimentale la un nivel calitativ superior. Cititorii interesaţi de o detaliere a acestor aspecte, precum şi de alte variante adecvate de alocare randomizată pot găsi mai multe informaţii în Torgerson şi Torgerson (2008).

Tabelul 4.2. *Soluţii optime de alocare randomizată a participanţilor în grupele experimentale*

Denumire	Descriere, avantaje şi dezavantaje
Randomizare simplă[5]	Este cea mai întâlnită formă de randomizare, uşor implementabilă prin apelul la soluţii software de generare a unor numere într-o manieră aleatorie, însoţită de decizia de a aloca participanţii în funcţie de anumite criterii (cei cărora li s-a alocat un număr par – în grupul experimental; cei cărora li s-a alocat un număr impar – în grupul de control). Avantajul acestei soluţii este că cel care face programările pentru evaluare sau evaluările (de eligibilitate) nu ştie în avans din ce grupă va face parte individul. Mai mult, dacă distribuirea se face de către un alt coleg, neimplicat în derularea experimentului, nu se va cunoaşte alocarea individului nici după finalizarea părţii de screening iniţial. Dezavantajul soluţiei este acela că poate conduce la grupuri inegale de subiecţi şi că, îndeosebi pentru un număr redus de participanţi/grupă, poate fi mai puţin eficientă în a controla variabilele străine. De asemenea, pentru studiile care necesită includerea unor participanţi aparţinând unui grup-ţintă greu de găsit (de exemplu, persoane care au trecut prin stadiul de moarte clinică), ar fi nevoie de un interval mai lung de timp pentru constituirea grupelor, fapt ce conduce la creşterea riscului ca în anumite perioade să fie selectaţi numai participanţi pentru un anumit grup (dacă calculatorul generează 5-6 numere consecutive indicative pentru acelaşi grup experimental).
Randomizarea restricţionată în perechi	Se caută numărul minim de participanţi (de exemplu, două persoane în cazul a două grupe, trei persoane în cazul a trei grupe) care sunt eligibili pentru a participa la studiu. Când acest număr minimal este atins, se apelează la numere

(*pairwise randomization*)	generate aleatoriu de către un program de calculator pentru a-i distribui în grupele aferente. Soluţia rezolvă problemele randomizării simple legate de inegalitatea numărului de subiecţi din grupe sau de secvenţialitatea/temporalitatea intervenţiei. Dezavantajul soluţiei constă în temporizarea aplicării intervenţiei (de exemplu, un participant eligibil trebui să aştepte până este identificată o altă persoană eligibilă în cazul existenţei a două grupe experimentale). În plus, soluţia are aceleaşi probleme precum randomizarea simplă, şi anume, riscul de a nu controla eficient toate variabilele străine în cazul unui număr redus de subiecţi per grup experimental.
Randomizarea restricţionată stratificată grupată	Înainte de randomizare, participanţii sunt împărţiţi după una sau două variabile străine considerate a fi esenţiale. De exemplu, într-un studiu care vizează eficienţa persuasivă a două tipuri de mesaje: cu accent pe aspecte cognitive (A) *vs* cu accent pe aspecte afective (B), randomizarea subiecţilor în grupele experimentale se face după ce în prealabil a fost măsurată complexitatea cognitivă a participanţilor, măsură care a condus la împărţirea prealabilă a participanţilor în indivizi cu complexitate cognitivă ridicată *vs* scăzută. Fiecare dintre cele două categorii a fost supusă apoi procesului de randomizare în grupe de câte patru persoane, astfel încât dacă grupului de patru persoane din categoria celor cu complexitate cognitivă ridicată le revin mesajele persuasive A, B, B, A, celuilalt grup de patru persoane din categoria celor cu complexitate cognitivă scăzută le revin mesajele B, A, A şi B. Calupul de patru litere (valabil pentru un design cu două modalităţi ale variabilei manipulate) îşi schimbă ordinea în mod aleatoriu (posibile fiind şase combinaţii: ABAB, AABB, BABA, BBAA, ABBA şi BAAB). Avantajul acestei soluţii este acela că permite controlul unor variabile străine importante inclusiv pentru un număr redus de subiecţi per celulă experimentală/grup şi că păstrează un echilibru numeric între grupele experimentale. Dezavantajele soluţiei constau în dificultatea de a controla mai mult decât o singură variabilă străină importantă (din motive de complexitate a designului) şi în faptul că ultima persoană din calupul de patru persoane este predictibilă din perspectiva apartenenţei sale la unul dintre cele două grupuri).
Randomizare adaptivă (minimizare)	Minimizarea nu este decât parţial o metodă de alocare randomizată a participanţilor în grupele experimentale, deoarece ea presupune alocarea iniţială (*random*) a jumătate dintre participanţii eligibili pentru studiu. De exemplu, dacă randomizarea are loc la nivel de şcoli, din cele opt şcoli contactate doar patru vor fi randomizate simplu în una din cele două grupe experimentale; celelalte patru

şcoli vor fi alocate pe bază raţională astfel încât să obţinem profiluri similare de comparat; de exemplu, dacă o şcoală din centrul oraşului, cu rată de promovabilitate ridicată la examenul de bacalaureat a fost distribuită iniţial prin randomizare în condiţia experimentală A, cealaltă şcoală localizată central şi având o rată de promovabilitate ridicată la bacalaureat, rămasă nedistribuită, va fi distribuită în condiţia experimentală B).
Pentru a reduce riscul de distorsiune în selecţie pentru şcolile rămase de distribuit după încheierea procesului de randomizare simplă, este preferabil ca algoritmii de potrivire să fie transpuşi într-un software pe calculator şi să se utilizeze o regulă bazată pe probabilităţi. De exemplu, şcoala potrivită pentru o anumită condiţie experimentală poate fi selectată de către calculator, în baza unei probabilităţi de 80%, nu a uneia de 100%. În asemenea condiţii, minimizarea poate fi văzută ca o metodă de randomizare adaptivă.

Pe scurt, randomizarea simplă şi randomizarea în perechi constituie cele mai bune soluţii de randomizare individuală a indivizilor atunci când avem un număr destul de ridicat de participanţi în fiecare condiţie experimentală, iar designul cercetării nu specifică vreun control special acordat unor variabile străine. În schimb, minimizarea este soluţia mai potrivită atunci când avem de a face cu un studiu efectuat pe un număr scăzut de persoane sau când randomizarea se aplică la nivel de clustere, în condiţiile în care găsirea de persoane eligibile pentru a participa la studiu necesită destul de mult timp. În sfârşit, pentru designurile factoriale parţial randomizate (*randomized blocked*), în care unele dintre VI sunt variabile etichetă, precum şi pentru cazurile în care se doreşte în mod explicit controlul unor variabile străine, iar numărul de participanţi la studiu este unul moderat, soluţia potrivită este aceea de a utiliza randomizarea stratificată, într-o variantă simplă sau grupată.

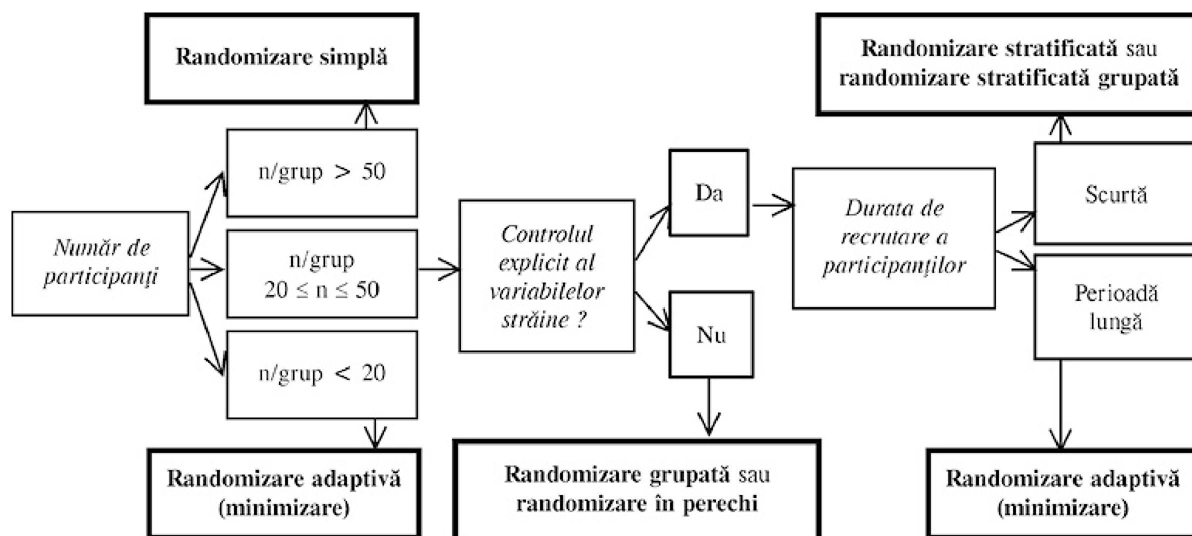


Figura 4.3. *Soluţii optime de alocare randomizată în funcţie de contextul cercetării*

Indiferent de modalitatea de randomizare aleasă, o randomizare de calitate ar trebui să includă metode de ascundere a rezultatului alocării în condiţiile experimentale (*allocation concealment*). Soluţia pentru a atinge acest deziderat vizează delegarea responsabilităţii randomizării. La ora actuală s-au dezvoltat servicii contra cost de randomizare a participanţilor în studiu (Torgerson şi Torgerson, 2008), însă este acceptabil şi să rugăm un coleg neimplicat în studiu să apeleze la generarea (extragerea) grupei din care face parte un participant declarat a fi eligibil, după ce a semnat, în prealabil, formularul de participare consimţită la studiu.

De asemenea este importantă mascarea cunoaşterii grupului în care a fost distribuită o anumită persoană de către cei care fac măsurătorile evoluţiei participanţilor la studiu în timp (pretest, interim, posttest, urmărire [*follow up*]). Astfel, aceştia nu vor fi influenţaţi în aprecierile lor asupra evoluţiei persoanei testate de cunoaşterea apartenenţei la unul sau altul dintre grupurile experimentale. În strânsă legătură cu acest punct se găseşte ideea mascării alocării randomizate şi a conceptului de „orb” (*blind*). Într-un studiu experimental pot fi identificaţi trei tipuri de actori: cei care oferă intervenţia (de exemplu, terapeuţii), cei care beneficiază de

intervenţie (de exemplu, clienţii) şi cei care evaluează progresele observate (de exemplu, asistenţii de cercetare). Aşadar putem vorbi în termeni populari de una dintre cele trei variante experimentale de orbire: simplu-orb, dublu-orb sau triplu-orb, în funcţie de câte categorii de actori nu cunosc cărui grup experimental îi aparţine o anumită persoană.

În domeniul psihologiei este puţin probabil ca toate cele trei categorii de actori implicaţi să poată fi supuşi condiţiei de orbire, deoarece este puţin probabil ca experimentatorul să nu cunoască la ce fel de intervenţie este supusă o anumită persoană. În schimb, asistenţii de cercetare care fac evaluarea progresului la nivelul VD şi uneori participanţii (prin includerea unui grup de control de tip placebo) se pot regăsi în situaţia de a nu cunoaşte grupul din care au făcut parte, acest lucru fiind menţinut până la finalul studiului, pentru a nu influenţa răspunsurile lor obţinute în perioada de urmărire (*follow-up*). Standardele CONSORT (Moher *et al.*, 2010) recomandă înlocuirea termenilor „orb", „dublu-orb" sau „triplu-orb" cu precizarea prezenţei sau absenţei condiţiei de orbire, iar în cazul din urmă, cu completarea categoriilor afectate de această condiţie şi prezentarea modalităţilor concrete în care s-a realizat „orbirea".

4.1.4. Controlul exercitat prin nivelul de comparaţie ales

Aşa cum am precizat încă de la începutul acestui capitol, ideea de control al unor variabile externe este una esenţială pentru succesul oricărui demers experimental. Un alt nivel de analiză privind controlul exercitat vizează nivelul comparativ, adică specificul grupului de control ales. Din acest punct de vedere, rezultatele grupului experimental pot fi comparate cu cele ale unui grup de control care poate lua forme diferite precum: (a)

control prin absenţa tratamentului; (b) control prin aplicarea unei intervenţii uzuale (ce se întâmplă de obicei); (c) control prin compararea cu intervenţia standard (cea mai bună soluţie existentă în acel moment); (d) control prin raportarea la o altă formă de intervenţie etc. Din acest punct de vedere Kazdin (2010) expune principalele tipuri de grup de control, sintetizate în tabelul 4.3.

Tabelul 4.3. *Tipuri de grup de control (lot martor) în funcţie de scopul cercetării*

Denumire şi descriere	**Utilitate (scop vizat)**	**Probleme**
Grup cu intervenţie absentă (*no treatment group*) Cei din grupul de control nu primesc niciun fel de intervenţie.	Se verifică dacă intervenţia are un efect. Este vorba însă despre un efect ce incorporează şi efectul determinat de alte aspecte, cum ar fi efectul placebo.	În anumite situaţii, absenţa unei intervenţii este discutabilă etic, îndeosebi dacă sunt avute în vedere cazuri medicale sau educaţionale în care participanţii la studiu au nevoie de o intervenţie. Se constată un număr mai mare de refuzuri de a participa la studii atunci când participanţii realizează că pot fi distribuiţi în grupul care nu primeşte niciun fel de tratament. Se observă un număr mai mare de refuzuri pentru a fi evaluaţi la nivel de posttest. Creşte potenţialul de risc la nivelul validităţii de construct a rezultatelor obţinute.
Grup cu intervenţie amânată (*waiting list group*) Cei din grupul de control primesc intervenţia doar la finalul derulării cercetării (posttest finalizat), uneori chiar după finalizarea perioadei de urmărire (*follow-up* finalizat).	Se verifică dacă intervenţia are un efect.	Poate fi aplicată doar în anumite situaţii (nu în situaţii acute/care necesită intervenţii imediate, deoarece ar ridica aceleaşi probleme etice ca şi în condiţiile absenţei intervenţiei). Dacă implementarea intervenţiei în grupul de control se realizează după evaluarea de la nivel de posttest, nu

		mai pot fi analizate efectele din perioada de urmărire (*follow-up*). Se observă un număr mai mare de refuzuri de participare la studiu, ce creşte cu cât perioada de aşteptare este una mai lungă. Durată mai lungă (dublă) de derulare a studiului.
Grup cu intervenţie nespecifică (*placebo* sau *nespecific intervention*) Cei din grupul de control primesc o intervenţie, pe care clienţii o percep ca fiind credibilă (utilă) pentru problema lor, dar care nu se bazează pe factorii activi vizaţi[6].	Se verifică dacă intervenţia vizată este responsabilă pentru efectul obţinut şi nu alte aspecte nespecifice precum (efectul placebo, alţi factori comuni, nespecifici etc.).	Obţinerea unor diferenţe semnificative între grupul experimental şi cel de control ar oferi doar siguranţa că nu factorii comuni/placebo sunt responsabili pentru efectul observat, dar nu vor oferi detalii despre care element specific este responsabil pentru efectul observat. Dificultatea de a dezvolta intervenţii care să fie percepute de participanţi drept acţiuni credibile pentru înlăturarea problemei constatate. Soluţie, uneori discutabilă din punct de vedere etic. Creşterea numărului de participanţi la studiu, deoarece mărimea efectului este mai mică dacă se compară tratamentul activ cu cel de tip placebo decât dacă se compară intervenţia activă cu „absenţa intervenţiei" sau cu „lista de aşteptare".
Grup cu intervenţie uzuală (*usual care*) Cei din grupul de control urmează procedurile existente la momentul respectiv.	Se verifică dacă intervenţia vizată este mai bună decât cea uzuală.	Preferabilă etic în faţa soluţiilor menţionate anterior deoarece şi grupul de control beneficiază de o intervenţie activă. Ceea ce se înţelege prin intervenţie uzuală poate varia de la o situaţie la alta, chiar pentru aceeaşi problemă.

		De exemplu, pentru reducerea absenteismului, o şcoală poate avea implementată o strategie clară bazată pe discuţii cu părinţii şi cu psihologul şcolar, în timp ce o altă şcoală apelează uzual la o discuţie între diriginte şi elev, urmată de informarea telefonică a părinţilor. Rezultatele studiului ar fi astfel dependente de locaţia unde s-ar derula studiul. Există un nivel mai scăzut de standardizare în implementarea (aplicarea) procedurilor implicate în intervenţia uzuală, fapt ce poate afecta încrederea în validitatea comparaţiei realizate. Creşterea numărului de participanţi comparativ cu cazul primelor două tipuri de grupuri de control („absenţa tratamentului sau aşteptarea”).
Grup cu intervenţie recomandată (*golden standard treatment group*) Cei din grupul de control urmează intervenţia care la momentul respectiv este recunoscută ca fiind cea mai întâlnită şi mai eficientă formă de acţiune pentru problema avută.	Se verifică dacă intervenţia vizată este mai bună, cel puţin egală sau măcar nu este inferioară intervenţiei considerate la momentul respectiv ca fiind cea mai eficientă sau cea mai răspândită soluţie.	Necesită cel mai ridicat număr de participanţi, comparativ cu celelalte strategii prezentate anterior, deoarece mărimile efectului privind diferenţele dintre cele două grupe au deseori valori scăzute. Este considerată a fi cea mai dezirabilă soluţie din punct de vedere etic.
Strategia aditivă Cazul este întâlnit atunci când grupul experimental beneficiază de o intervenţie activă suplimentară (de	Se verifică dacă intervenţia adăugată (combinată) poate aduce un beneficiu suplimentar	Necesită un număr ridicat de participanţi, deoarece câştigul adăugat de un eventual al doilea element activ nu este la fel de mare ca şi diferenţa existentă atunci când

exemplu, terapie, în plus faţă de medicaţie), iar performanţa se compară cu un grup de „control” care primeşte doar medicaţie.	(incremental) faţă de implementarea unei intervenţii singulare.	comparăm cazul aplicării unui element activ cu absenţa oricărui element activ. Necesită mai multe resurse pentru a putea fi implementată. Poate fi implementată doar dacă în prealabil au fost colectate dovezi cu privire la eficienţa intervenţiilor singulare sau dacă se aşteaptă ca una dintre strategii să potenţeze efectul celeilalte (să joace un rol moderator).
Strategia derivativă În acest caz nu se poate vorbi propriu-zis de un grup de control, deoarece ambele grupe primesc: (a) fie acelaşi tip de intervenţie în gradiente diferite (expunere la stimuli timp de 10 minute sau 30 de minute); (b) fie forme variate ale aceluiaşi tip de intervenţie (terapie cognitiv-comportamentală centrată pe combaterea cogniţiilor iraţionale asociate toleranţei scăzute la frustrare *vs* asociate autodeprecierii)	Se verifică ce modificări în intervenţie pot conduce la creşterea eficienţei acesteia. Se analizează care componente din cadrul intervenţiei sunt responsabile pentru efectul observat. Cum poate fi optimizată intervenţia?	În absenţa introducerii unui grup de control[7], acest tip de design necesită cel mai mare număr de participanţi pentru a asigura o putere statistică optimă în testarea ipotezelor. Are avantajul de a trece dincolo de simpla discuţie despre prezenţa sau absenţa unei relaţii de tip cauză-efect spre a oferi explicaţii cu privire mecanismele cauzale implicate (mecanismele schimbării). Desigur, poate fi implementată doar după ce există deja dovezi cu privire la eficienţa intervenţiei analizate, la nivel global (molar).

Întreaga discuţie menţionată anterior porneşte de la premisa necesităţii existenţei unui grup de control, pentru a putea analiza dacă eventualele modificări observate la nivelul grupului experimental sunt cauzate de variabila independentă manipulată. În absenţa unui asemenea grup comparativ, nu vom putea determina dacă modificările sunt cauzate de aceşti factori sau de alţi factori care ar putea intra în categoria variabilelor confundate (vezi şi secţiunea destinată riscurilor la nivelul validităţii

interne din capitolul 3). De exemplu, stabilirea eficienţei terapiei cognitiv-comportamentale în ameliorarea simptomatologiei asociate enurezisului prin apelul la un design cu un singur grup pretest-posttest, fără grup de control, nu poate fi stabilită deoarece eventualele ameliorări observate pot fi cauzate atât de intervenţia în sine, cât şi de procesul normal de maturare (reducerea simptomatologiei pe măsura înaintării în vârstă).

La fel de important de reţinut este şi influenţa diferenţiată pe care o are tipul de grup de control asupra validităţii concluziilor unui studiu. De exemplu, alipirea unui grup de control de tipul absenţei tratamentului va permite eliminarea explicaţiei alternative conform căreia maturarea ar fi elementul responsabil, deoarece atât grupul experimental, cât şi cel de control trec prin aceleaşi condiţii. Însă un asemenea design nu ar putea permite unui autor să afirme că reducerea cogniţiilor iraţionale ar fi responsabilă pentru ameliorarea simptomatologiei deoarece nu a existat un grup de control activ (placebo, intervenţie uzuală) care să elimine impactul factorilor comuni (nespecifici). De asemenea, nu s-ar putea valida nici afirmaţii cum ar fi: (a) terapia respectivă este cea mai eficientă soluţie pentru combaterea enurezisului (în lipsa unor grupuri de control active precum grupurile de intervenţie recomandată); (b) terapia ar trebui utilizată ca tratament adjuvant la medicaţie (atât timp cât nu există dovezi că acţiunea lor combinată ar fi mai eficientă decât utilizarea exclusivă a tratamentului medicamentos).

4.2. Studiile randomizate controlate

În situaţia în care studiile experimentale sunt utilizate în contexte specifice cercetării pentru a răspunde la întrebări precum: (a) „este

intervenţia X una eficientă?”; (b) „care componentă a intervenţiei X este responsabilă pentru eficienţa intervenţiei (care este mecanismul schimbării)?”; (c) „este intervenţia X la fel de eficientă în cazul şcolarilor mici precum în cazul liceenilor?” etc., soluţia metodologică potrivită este aceea de a realiza studii randomizate controlate (*randomized controlled trial*)[8].

Deşi cele mai multe studii randomizate controlate sunt efectuate în domeniul ştiinţelor medicale, aplicabilitatea lor este extinsă la o sferă mult mai largă de domenii, precum domeniul clinic/al psihoterapiei dominant în psihologie, cel educaţional, domeniul juridic, domeniul organizaţional, domeniul politicilor publice etc. Din perspectiva paradigmei oricărei ştiinţe validate ştiinţific (aspectul principal analizat în lucrarea de faţă), aceste studii sunt esenţiale, deoarece au cel mai ridicat nivel de validitate internă, prin urmare concluziile acestor studii prezintă cel mai scăzut risc de a fi distorsionate.

În titulatura acestui gen de studii se regăsesc sintagmele „randomizate”, respectiv „controlate”, acestea fiind specifice oricărui studiu experimental. În consecinţă, aspectele discutate în secţiunile anterioare ale acestui capitol se aplică şi pentru acest tip particular de studii experimentale. În plus, o serie de amănunte specifice studiilor randomizate controlate au fost conceptualizate (vezi standardele CONSORT, Schulz *et al.*, 2010; Moher *et al.*, 2010) şi merită discutate în continuare, în ideea facilitării implementării acestui gen de studii în spaţiul românesc.

Prima idee care ar trebui reţinută în legătură cu studiile randomizate controlate este aceea că sub umbrela acestei expresii intră o varietate de studii, a căror clasificare sintetică după criterii variate se regăseşte în figura 4.4.

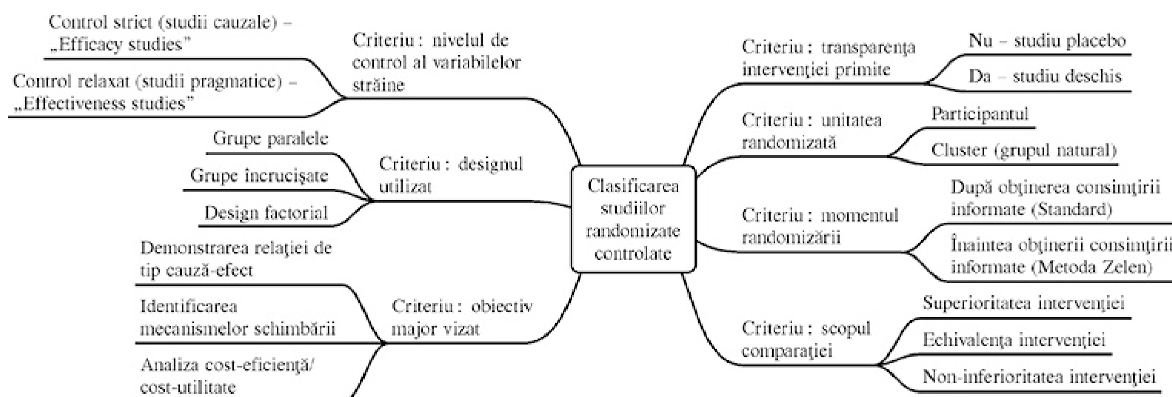


Figura 4.4. *Clasificarea studiilor randomizate controlate după criterii variate*

O scurtă descriere a elementelor menţionate în figura 4.4 se regăseşte în tabelul următor, cu menţiunea că o detaliere a diferenţelor cele mai importante dintre diferitele tipuri de studii randomizate este prezentată în secţiunea 4.2.1, în care sunt descrise pe larg etapele unui prototip de studiu randomizat controlat.

Tabelul 4.4. *Descrierea succintă a aspectelor menţionate în figura 4.4*

Criteriu: nivelul de control al variabilelor străine • Studiile cauzale (*efficacy studies*) au drept caracteristică un control strict al tuturor variabilelor străine la toate nivelurile (în selectarea participanţilor eligibili în cadrul studiilor; în monitorizarea aplicării într-o manieră integră a intervenţiei; în selectarea variabilelor dependente primare etc.). Prin aceste măsuri asiguratorii se are în vedere obţinerea unei validităţi interne ridicate pentru concluziile studiului, fapt ce permite inclusiv identificarea unor relaţii de tip cauză-efect şi/sau a mecanismelor schimbării implicate. • Studiile pragmatice (*effectiveness studies*) au drept caracteristică un control mai relaxat al unor variabile importante pentru studiu, în ideea replicării situaţiilor din realitate. Astfel, în studiile pragmatice, criteriile de eligibilitate pentru selectarea participanţilor sunt mai puţin stricte şi se acceptă devieri faţă de cele specificate în protocolul de implementare a intervenţiei, iar variabilele dependente selectate sunt cele importante/relevante pentru societate/individ (de obicei, aspecte comportamentale). Prin aceste măsuri asiguratorii se are în vedere îndeosebi obţinerea unei validităţi externe (ecologice) ridicate. Dat fiind specificul celor două tipuri de studii, studiile cauzale ar trebui să preceadă temporal studiile pragmatice, acestea din urmă având sens doar dacă în primele se identifică o relaţie de tip cauză-efect. Mai multe detalii despre diferenţele dintre aceste tipuri de studii sunt prezentate în secţiunea 4.2.1.

Criteriu: designul utilizat

• Designul de bază cu grupe paralele (*parallel groups* sau *parallel arms*) reprezintă, de departe, cel mai frecvent design întâlnit în studiile randomizate controlate, constând în compararea a două sau mai multe grupe diferenţiate la nivelul modalităţilor variabilei independente (de exemplu, compararea terapiei raţional-emotive, cu un grup de control de tipul listei de aşteptare privind ameliorarea simptomatologiei depresive). Modelul standard de lucru într-un design cu grupe paralele presupune: (a) verificarea eligibilităţii participanţilor la studiu, cu ocazia evaluării la nivel de pretest; (b) randomizarea participanţilor în grupe şi aplicarea intervenţiei (pentru grupul experimental); (c) evaluarea şi compararea efectelor aplicării, respectiv neaplicării intervenţiei la momentul postintervenţie; (d) evaluarea şi compararea efectelor intervenţiei la finalul perioadei de urmărire (*follow-up*).

• Designul de bază cu grupe încrucişate (*cross-over design/groups/arms*) reprezintă o extensie a modelului cu grupe paralele, prin faptul că după realizarea primilor trei paşi (după măsurarea şi compararea efectelor observate la grupul experimental, respectiv cel de control, de la momentul postintervenţie), cele două grupe îşi schimbă rolul, astfel încât intervenţia activă este aplicată asupra fostului grup de control, în timp ce fostul grup experimental devine grupul de aşteptare (de control). Acest design este rar întâlnit în domeniul ştiinţelor sociale din cauza efectului de reportare (*carry-over*), a prelungirii duratei cercetării sau a renunţării de a studia efectele la finalul perioadei de urmărire (*follow-up*).

• Designul factorial (*factorial design*) reprezintă cazul în care grupele studiului sunt alcătuite în baza combinării a două (cel mai adesea) sau mai multe variabile independente. Un exemplu în acest sens ar fi un studiu în care se urmăreşte dacă un program de consiliere psihologică ar fi eficient în a-i determina pe adulţi să renunţe la fumat, în contextul în care se doreşte a se ţine seama şi de eficienţa deja dovedită a plasturilor cu conţinut de nicotină. Din încrucişarea celor două variabile independente rezultă patru celule experimentale posibile: participanţi care vor primi atât consiliere psihologică, cât şi tratament farmaceutic prin aplicarea de plasturi; participanţi care vor beneficia doar de consiliere psihologică; participanţi care vor primi doar tratament farmaceutic şi participanţi aflaţi în grupul de control (fără intervenţie). Avantajul designului factorial faţă de celelalte tipuri de design constă în posibilitatea de a studia efectul de interacţiune dintre cele două tipuri de intervenţii. Dezavantajul constă în consumul ridicat de resurse şi în necesitatea identificării mai întâi a două forme diferite de intervenţie posibile (eficiente), aspecte care fac ca acest tip de design să fie mai rar întâlnit comparativ cu designul bazal cu grupe paralele.

Criteriu: obiectivul major al studiului

• Demonstrarea unei relaţii de tip cauză-efect este cel mai frecvent obiectiv întâlnit în cadrul studiilor randomizate controlate, primele întrebări fiind adesea: „Este această intervenţie eficientă? Se poate spune că există o relaţie de tip cauză-efect între expunerea la X şi apariţia efectului Y?”. Acest tip de obiectiv poate fi atins prin acordarea unei atenţii deosebite aspectelor

de validitate internă ale unei cercetări, de aceea el este atins prin studii cauzale (vezi primul criteriu).

- Identificarea mecanismelor schimbării reprezintă un alt obiectiv important ce este adresat prin intermediul studiilor randomizate controlate, această întrebare fiind pusă după demonstrarea unei relaţii de tip cauză-efect. Prin intermediul acestui obiectiv se vizează răspunsuri la întrebări de genul: „Ce componentă a intervenţiei/ce anume din intervenţie a condus la apariţia unei relaţii de tip cauză-efect?”, „Care este elementul activ?”, „Ce aspecte din cadrul intervenţiei sunt responsabile pentru apariţia unei relaţii de tip cauză-efect?”. Acest tip de obiectiv poate fi atins doar prin intermediul unei forme particulare de studii cauzale (*efficacy studies*), pe care am etichetat-o în capitolul 3 ca cercetare de tip cauzal-explicativ.
- Din perspectiva unei abordări validate ştiinţific, după demonstrarea unei relaţii de tip cauză-efect (este intervenţia eficientă?), după identificarea mecanismelor schimbării responsabile (de ce intervenţia este una eficientă?) se poate stabili dacă intervenţia are un raport cost-eficienţă sau cost-utilitate pozitiv (cât de rentabilă este intervenţia?). Deşi acest tip de întrebare poate fi abordat prin designuri de cercetare multiple (inclusiv de tip non-experimental), cel mai potrivit design este cel experimental (preferabile de această dată fiind studiile pragmatice, pentru a evidenţia costurile şi efectele obţinute într-un context cât mai apropiat de realitate, în condiţiile menţinerii unui control asupra unor potenţiale variabile străine).

Criteriu: transparenţa intervenţiei primite

- Studiile care includ condiţia placebo (*placebo/sham studies*) reprezintă standardul de lucru în domeniul ştiinţelor medicale. Acest lucru presupune mascarea intervenţiei active, prin intermediul unor intervenţii nespecifice (neactive), cât mai apropiate de intervenţia activă pentru a elimina anumite distorsiuni posibile care apar în contextul în care participanţii cunosc faptul că nu sunt incluşi în grupul experimental activ (vezi efectul Hawthorne sau efectele de demoralizare cu consecinţe negative, inclusiv la nivelul creşterii procentului de renunţare la studiu din rândul participanţilor incluşi în grupele inactive). Pentru a potenţa efectele pozitive, studiile care ţin seama de efectul placebo ar trebui să apeleze la o măsură suplimentară de siguranţă. Ar fi indicat nu numai ca participantul să nu cunoască exact grupul din care face parte, ci şi cei care implementează studiul (aplică intervenţia) sau efectuează evaluările pentru a detecta prezenţa sau absenţa unui efect să nu ştie din ce grup face parte un participant. Dacă mascarea grupului de apartenenţă este posibilă pentru toate cele trei categorii de actori vorbim despre un design triplu-orb, dacă mascarea este posibilă doar pentru două categorii de actori avem un design dublu-orb. În contextele specifice domeniului psihologiei este destul de greu să se implementeze studii de tip placebo (design orb pentru participanţi) sau designuri cu un nivel de mascare mai complexă (deoarece, spre exemplu, psihoterapeutul sau trainerul trebuie să ştie ce fel de intervenţie să aplice asupra unei persoane anume). Totuşi, în cazul studiilor din sfera psihologiei este posibilă mascarea pentru o categorie de actori, aceia care realizează evaluările, măsurând efectele produse de intervenţie la diferite intervale de timp. Acest lucru este necesar mai ales dacă măsurile colectate au un caracter subiectiv (observaţii), caz în care cunoaşterea grupului căruia îi aparţine o

anumită persoană ar putea conduce la aprecieri distorsionate, care să favorizeze efectul intervenţiei.

• Studiile deschise (transparente) (*open studies*) reprezintă acele situaţii experimentale în care toţi participanţii implicaţi activ (participanţi, traineri, evaluatori) cunosc cărui grup îi aparţine fiecare persoană inclusă în studiu. În acest gen de studii, pentru a reduce eventualele distorsiuni legate de menţinerea unora dintre participanţi în grupe inactive, ar fi preferabil să se utilizeze grupe de control active (intervenţii uzuale, intervenţii recomandate/comparative). O altă soluţie recomandabilă ar fi aceea de a selecta drept variabile dependente principale măsuri fiziologice sau comportamentale care să fie mai puţin susceptibile la distorsiune comparativ cu măsurile de tip autoraportare (*self-report*) sau hetero-raportare (*peer-report*), mai ales dacă grupul de control este unul de tip inactiv (fără intervenţie sau listă de aşteptare). În sfârşit, o altă recomandare ar fi să se ia în considerare preferinţele participanţilor şi să se analizeze impactul acestora în rezultatele finale (efectul observat la participanţii distribuiţi în grupul A la cei care preferau să fie distribuiţi în această intervenţie comparativ cu efectul participanţilor din grupul A care ar fi preferat să fie distribuiţi în grupul B).

Criteriu: unitatea supusă randomizării

• Individul reprezintă soluţia standard de lucru în studiile randomizate controlate. Astfel, randomizarea participanţilor la studiu în condiţiile experimentale se face individual, aceasta fiind soluţia preferabilă ori de câte ori este posibil acest lucru, pentru a asigura un control strict al variabilelor străine care ţin de caracteristicile indivizilor incluşi în studiu. Detalii suplimentare despre acest tip de studiu vor fi prezentate în continuarea acestei secţiuni.

• Grupul natural reprezintă soluţia alternativă de alocarea randomizată a indivizilor în condiţiile experimentale, atunci când nu este fezabilă randomizarea la nivel individual. Acest nivel de randomizare poartă denumirea de randomizare cluster. Pentru a fi la fel de eficientă precum randomizarea la nivel de individ, randomizarea la nivel de cluster trebuie să respecte câteva condiţii, discutate în secţiunea 4.2.2. Acest gen de randomizare este uneori singura soluţie (de exemplu, în medii organizaţionale sau şcolare unde s-ar putea obţine cu greu acceptul de a redistribui echipele de muncă/clasele de elevi). În acest context, elementele randomizate vor fi clasele/echipele din care participanţii fac parte în mod natural. Avantajul acestui tip de randomizare în condiţiile lucrului în medii cu grupuri naturale gata formate (departamente, clase etc.) este acela de a împiedica diseminarea informaţiei din grupurile active şi de a reduce efectele de tip Hawthorne sau cel de demoralizare/demotivare). Detalii despre randomizarea la nivel de cluster găsiţi în secţiunea 4.2.2.

Criteriu: momentul randomizării

• Modalitatea standard de lucru într-un studiu randomizat controlat este aceea de a aloca aleatoriu participanţii eligibili în grupele experimentale după ce aceştia au citit în prealabil informaţii despre studiu şi au consimţit să participe la studiu, fără a avea dreptul de a alege grupa experimentală în care vor fi distribuiţi. Această variantă este atât de frecvent întâlnită, încât nici nu are o titulatură specifică, fiind, de altfel, varianta preferabilă şi din punct de vedere etic.

Problema identificată cu acest tip de design este aceea că el nu ia în considerare impactul preferinţelor participanţilor (deşi în general efectul unei intervenţii tinde să fie mai ridicat în cazul pacienţilor atraşi de intervenţia respectivă şi mai scăzut în cazul celor care au o opinie nefavorabilă faţă de intervenţia respectivă). Alocarea randomizată nu rezolvă întotdeauna această problemă. De exemplu, în cazul reducerii durerii postoperatorii este posibil ca majoritatea participanţilor să prefere apelul la farmacoterapie (din motiv de timp sau confort) decât să fie distribuiţi într-un grup de psihoterapie, iar această inegalitate în preferinţe poate afecta validitatea internă a rezultatelor obţinute.

• Metoda alternativă de lucru, cunoscută sub denumirea de metoda Zelen (*apud* Torgerson şi Torgerson, 2008), presupune randomizarea participanţilor în grupele studiului înainte de a cere acceptul acestora de a participa la studiu. Pentru a considera această soluţie acceptabilă din punct de vedere etic (în contextul în care participanţii nu cunosc faptul că au fost alocaţi aleatoriu anterior, fără acceptul lor), ea poate fi utilizată în contexte în care grupul de control primeşte un tratament activ (cel puţin intervenţia uzuală sau intervenţia standard/recomandată). În varianta simplă a metodei Zelen, doar participanţii incluşi în grupul care primeşte intervenţia inovativă sunt chestionaţi cu privire la preferinţa lor, iar cei care se declară împotriva tratamentului inovativ sunt incluşi în grupul de control (beneficiar al intervenţiei recomandate/uzuale). În varianta dublă, prin metoda Zelen sunt chestionate ambele categorii de indivizi, atât cei alocaţi grupului experimental, cât şi cei alocaţi grupului de control, urmând ca participanţii care ar prefera celălalt tratament existent să fie
realocaţi conform preferinţei lor. Un aspect tehnic important legat de aceste proceduri, pentru a menţine studiul în categoria studiilor randomizate, este faptul că analiza statistică a rezultatelor se face ţinând cont de recomandările strategiei intenţiei de a trata (*intention to treat*) (vezi secţiunea 4.2.3). În cazul de faţă, acest lucru înseamnă că rezultatele celor două grupe sunt analizate pornind de la compoziţia iniţială a grupelor în urma randomizării. De exemplu, indivizii care au fost realocaţi pentru a primi intervenţia B datorită preferinţei lor, vor fi analizaţi ca făcând parte din grupul A, unde fuseseră iniţial repartizaţi aleatoriu, în baza strategiei bazată pe intenţia de a trata. Desigur, o asemenea soluţie este posibilă dacă numărul celor realocaţi în total între grupe nu este unul ridicat (sub 10-15%), altfel fiind afectate concluziile finale din cauza mixării intervenţiilor. O alternativă posibilă în analiza datelor provenite dintr-un studiu ce apelează la metoda Zelen este analiza standard a datelor, dublată de efectuarea unor analize suplimentare cu referire la impactul pe care preferinţa participanţilor pentru o anumită intervenţie îl joacă asupra rezultatelor finale.

Criteriu: scopul comparaţiei

• Demonstrarea superiorităţii unei intervenţii (grupul experimental) în faţa grupului de control este varianta tipică întâlnită în studiile randomizate controlate, fiind extrem de răspândită. Scopul ei este de a urmări în ce măsură o intervenţie poate fi considerată a fi semnificativ mai eficientă decât alte soluţii supuse comparării. Absenţa unui rezultat semnificativ statistic în urma

comparării unor intervenţii, fie ele A şi B, nu înseamnă automat că cele două intervenţii sunt echivalente sau că se află într-o relaţie de non-inferioritate una faţă de cealaltă.

• Demonstrarea echivalenţei dintre două intervenţii se întâlneşte atunci când o intervenţie nouă este comparată cu o intervenţie a cărei eficienţă a fost demonstrată, pentru a arăta că ambele au un nivel similar de eficienţă. Demonstrarea non-inferiorităţii unei intervenţii este utilizată atunci când se doreşte a se arăta că o intervenţie nouă nu este mai puţin eficientă (inferioară) unei intervenţii consacrate pentru problema respectivă. În ambele cazuri se defineşte un interval marginal (o diferenţă Δ considerată ca fiind maximumul distanţei posibile dintre două intervenţii care nu are un impact practic [clinic] – de exemplu, mărimea efectului de intensitate mică). Spre ilustrare, două intervenţii sunt considerate echivalente în tratamentul depresiei dacă diferenţele sesizate nu sunt semnificative statistic şi dacă ele se regăsesc în intervalul stabilit anterior (de exemplu, o diferenţă de ±2 puncte între mediile celor două grupe comparate la o probă care evaluează nivelul depresiei). Transpus în termeni de non-inferioritate, am spune despre o intervenţie menită să combată depresia că ea este non-inferioară alteia consacrate, dacă diferenţa dintre medii nu este semnificativă statistic şi dacă nivelul mediu al depresie din grupa care a beneficiat de intervenţia experimentală nu este mai mic de 2 puncte faţă de nivelul mediu al depresiei din grupul care a primit intervenţia standard (recomandată). Aşadar, echivalenţa se transpune în termenii unei ipoteze bilaterale, în timp ce non-inferioritatea este specifică unor ipoteze unilaterale. În cazul demonstrării echivalenţei/non-inferiorităţii unei intervenţii noi, comparativ cu intervenţia standard, noua intervenţie poate deveni prima opţiune/soluţie dacă ea posedă alte caracteristici superioare (cum ar fi costuri mai reduse, acceptabilitate mai ridicată din partea beneficiarilor, durată mai redusă de implementare, aria de diseminare mai largă etc.). Mai multe detalii privind implicaţiile metodologice pentru studii randomizate controlate ce vizează demonstrarea echivalenţei/non-inferiorităţii unei intervenţii faţă de un alt tratament activ (recomandat) sunt disponibile în Piaggio *et al.* (2006).

Deşi numărul posibilităţilor de a combina diferitele caracteristici ale studiilor randomizate controlate este unul nelimitat, cel mai întâlnit design de acest gen în domeniul psihologiei este cel al unui (a) studiu cauzal (*efficacy study*) (b) cu un design bazal cu două grupe paralele, (c) deschis sau simplu-orb (prin mascarea apartenenţei la grupe în faţa asistenţilor de cercetare care măsoară evoluţia variabilelor dependente), (d) ce urmăreşte demonstrarea unei relaţii de tip cauză-efect, aspect căruia i se adaugă uneori identificarea mecanismelor schimbării (e) şi în care alocarea randomizată a participanţilor în grupe se face la nivel individual, (f)

ulterior obţinerii consimţirii informate din partea participanţilor. Din aceste motive, în secţiunea următoare a acestui capitol, vom detalia etapele realizării unui asemenea studiu-prototip, oferind însă specificaţii paralele suplimentare cu rol comparativ pentru situaţia studiilor pragmatice (*effectiveness*).

4.2.1. Etapele derulării unui studiu randomizat controlat

Pentru a structura conţinutul acestei secţiuni, vom apela la reprezentarea grafică a unui studiu randomizat controlat, conform standardelor CONSORT (Moher *et al.*, 2010), care are în vedere etapele acestui proces: (a) recrutarea şi selectarea participanţilor (*enrollment*); (b) alocarea randomizată a participanţilor şi derularea intervenţiilor; (c) urmărirea evoluţiei acestora (*follow-up*); (d) analiza statistică a rezultatelor.

4.2.1.1. Etapa de recrutare şi selecţie a participanţilor

Aparent, această etapă pare una mai facilă, cel puţin comparativ cu celelalte etape implicate într-un studiu randomizat controlat. În realitate, însă, nu mai puţin din 54% dintre studiile de acest gen au apelat la o extindere a perioadei de recrutare pentru a-şi atinge numărul dorit de participanţi eligibili, iar 45% dintre studii s-au derulat cu un număr de participanţi sub nivelul a 80% din numărul estimat (dorit) iniţial (McDonald *et al.*, 2006)

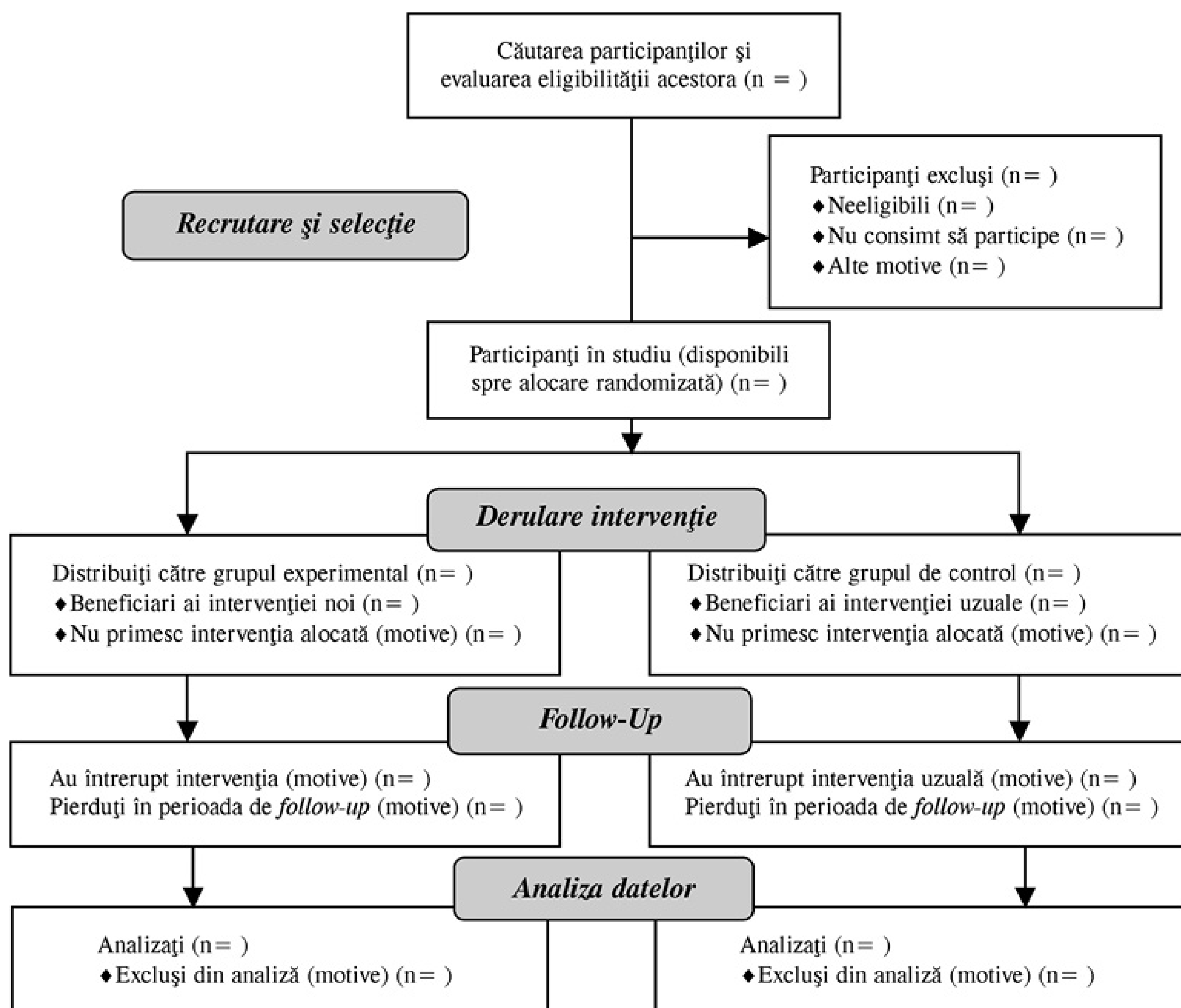


Figura 4.5. *Diagrama etapelor existente în implementarea unui studiu randomizat controlat cu două grupe paralele, cu un grup de control activ – intervenţie uzuală*

O recrutare de succes este aceea care reuşeşte să strângă în intervalul de timp prestabilit numărul de persoane eligibile pentru a primi intervenţia, aşa cum s-a estimat iniţial pentru a asigura o putere statistică optimă de a detecta efectul intervenţiei asupra variabilei dependente principale (*primary outcome*). În demersul de recrutare se poate apela la orice mijloc de promovare a studiului, de la publicitatea efectuată în mass-media locală la extinderea numărului de locaţii (instituţii colaboratoare care implementează studiul) şi la oferirea de stimulente financiare care să fie acceptabile din punctul de vedere al comitetului de

etică a cercetării. De exemplu, în SUA, plata acordată direct participanţilor eligibili pentru a fi incluşi în studiu este considerată adecvată doar dacă se rezumă la a acoperi unele cheltuieli ale acestora (cum ar fi cele de deplasare), avantajul constând mai degrabă în oferirea intervenţiei în regim gratuit. Recomandabil ar fi însă să fie motivaţi financiari alţi specialişti (de exemplu, profesori, medici etc.) care să ajute la identificarea cazurilor potenţial eligibile (Torgerson şi Torgerson, 2008). O altă soluţie ingenioasă, de interes pentru organizaţiile orientate spre profit, ar fi aceea de a scrie direct angajatorilor propunându-li-se acestora realizarea de parteneriate (aceştia să îndemne/cointereseze angajaţii să participa la studii) dacă problema vizată va duce la o reducere a cheltuielilor de sănătate (creştere a profitului), cum ar fi cazul programelor destinate reducerii fumatului sau reducerii stresului la locul de muncă (Lando, 1982).

Participanţii identificaţi ca fiind potenţial eligibili sunt evaluaţi iniţial pentru a determina dacă ei îndeplinesc într-adevăr criteriile de includere în studiu, respectiv dacă nu intră sub incidenţa unor criterii de excludere. Aceste criterii sunt stabilite *a priori*, fiind mai restrictive în cazul studiilor cauzale decât în cazul celor pragmatice. Desigur, indicat ar fi să fie luate în considerare doar acele criterii de includere/excludere care au relevanţă pentru obiectivul cercetării, astfel încât să nu fie afectată validitatea internă (pe de o parte – vezi cazul studiilor cauzale), dar şi să fie potenţată validitatea externă a studiului (capacitatea de generalizare a concluziilor – vezi cazul studiilor pragmatice). Un alt rol al acestei evaluări iniţiale este de a colecta rezultatele participanţilor la studiu la momentul preintervenţiei (pretest), ce vor fi utilizate ulterior în analizele statistice.

Un aspect important de menţionat cu privire la această etapă este aspectul temporal. Astfel, în multe cazuri, etapa de recrutare şi selecţie se

realizează integral înaintea începerii studiului, aşa cum este prezentată şi situaţia din figura 4.5. Totuşi, în studiile în care este nevoie de mulţi participanţi sau în care participanţii sunt greu de găsit, o soluţie posibilă este aceea de a suprapune etapa de recrutare cu etapele ulterioare de derulare a studiului, caz în care vorbim de o recrutare secvenţială (de exemplu, este posibil ca la momentul recrutării ultimilor participanţi necesari pentru studiu, primii indivizi recrutaţi să fie deja în perioada de urmărire [*follow-up*]). Ceea ce este important în aceste cazuri este ca intervenţia să aibă o durată şi activităţi standard şi să existe un procent relativ egal de persoane distribuite în grupul experimental, respectiv de control, într-o anumită etapă, astfel încât să fie controlate eventuale efecte sezoniere sau efecte ce ţin de instrumentare (variaţii în familiarizarea cu sarcina).

Caseta 4.4. *Ilustrarea unor criterii de includere/excludere în cazul ipotetic al unui studiu randomizat controlat care urmărea impactul psihoterapiei în combaterea depresiei majore*

A. Studiu cauzal

Participanţii eligibili sunt cei: (a) cu vârste cuprinse între 18 şi 65 de ani; (b) care au fost diagnosticaţi ca având tulburare depresivă majoră conform standardelor DSM IV; (c) care nu prezintă alte comorbidităţi psihiatrice; (d) care obţin un scor de cel puţin 20 la proba de depresie Beck; (e) care nu sunt implicaţi în vreo altă formă de tratament pentru ameliorarea simptomatologiei depresive; (f) care se află la primul episod depresiv diagnosticat.

B. Studiu pragmatic

Participanţii eligibili sunt cei: (a) cu vârste cuprinse între 18 şi 65 de ani; (b) diagnosticaţi cu tulburare depresivă majoră conform standardelor DSM IV; (c) care nu sunt implicaţi în vreo altă formă de tratament pentru ameliorarea simptomatologiei depresive.

Participanţii care în urma evaluării iniţiale îndeplinesc condiţiile de eligibilitate sunt solicitaţi să semneze un formular de consimţire informată

pentru a putea fi incluşi în studiu, acest procedeu fiind necesar pentru a îndeplini standardele etice în derularea cercetărilor.

Deşi formatul unui formular de consimţire informată poate varia de la un studiu la altul, în conţinutul ideatic al acestuia ar trebui să se regăsească, indiferent de tipul studiului (experimental, non-experimental etc.), descrieri ale: (a) obiectivului cercetării; (b) duratei studiului, în special ale duratei de implicare activă a participantului la studiu; (c) procedurilor şi sarcinilor care vor trebui urmate (inclusiv ale faptului că participanţii vor fi alocaţi aleatoriu în grupele experimentale, cu descrierea specificului fiecărui tip de intervenţie şi a evaluărilor care vor fi făcute); (d) potenţialelor riscuri existente prin expunerea la intervenţia oferită (implicarea în cercetare); (e) eventualelor beneficii rezultate din participarea la studiu (altele decât stimulentele financiare/academice); (f) stimulentelor financiare/academice (de exemplu, credite suplimentare, scutiri de sarcini, bonus în notare), dacă acestea vor fi acordate, cu precizarea explicită a faptului că acestea nu sunt condiţionate de completarea integrală a sarcinilor prevăzute în studiu/renunţarea pe parcurs la studiu; (g) modului în care va fi asigurată confidenţialitatea datelor şi, dacă este cazul, solicitarea permisiunii de a utiliza unele materiale colectate (fotografii, înregistrări audio) în scop de cercetare; (h) drepturilor participantului la studiu (de a completa doar sarcinile/porţiuni din sarcini în care consimte voluntar, de a se retrage pe parcursul studiului fără a fi penalizat sau a pierde vreun beneficiu câştigat până atunci; de a cere asistenţă gratuită în cazul în care procedurile experimentale i-au creat un disconfort/o problemă de sănătate); (i) datelor de contact (telefon, e-mail, adresă poştală) ale coordonatorului studiului sau ale persoanei desemnate să răspundă la orice problemă/întrebare pe care o au, precum şi a datelor de contact ale unei instanţe superioare, în cazul în care nu sunt

mulţumiţi de modul în care le-au fost soluţionate cererile de către coordonatorului studiului/persoana de contact desemnată; (j) declaraţiei-tip a participantului, cu spaţiu pentru a trece olograf numele, semnătura şi data, în care se specifică faptul că formularul a fost citit, iar participantul este de acord cu aspectele specificate. Acest formular este redactat în două exemplare, câte unul de fiecare parte (participant, respectiv echipa de cercetare), fiind de asemenea semnat de către coordonatorul studiului.

Kazdin (2010), unul dintre cei mai mari specialişti în domeniul metodologiei cercetării în psihologie clinică şi psihoterapie, recomandă ca formularul de consimţire informată să aibă următoarele secţiuni/subtitluri: (a) privire de ansamblu asupra studiului (*overview*); (b) descrierea procedurilor studiului; (c) riscuri; (d) beneficii; (e) stimulente; (f) confidenţialitate; (g) intervenţii alternative disponibile către client (în afara studiului); (h) acordul voluntar – un paragraf cu acordul voluntar al participantului la studiu; (i) date de contact pentru întrebări şi informaţii suplimentare; (j) semnături.

Un exemplu concret de formular de consimţire informată, folosit într-un studiu clinic controlat realizat în România, este prezentat în caseta 4.5.

Caseta 4.5. *Un model de formular de consimţire informată utilizat într-un studiu randomizat*

ACORD DE PARTICIPARE LA CERCETARE

NUMELE CERCETĂRII (STUDIULUI)

Rolul psihoterapiei cognitiv comportamentale în tratamentul chirurgical parodontal: un studiu clinic controlat (PSIDENT).

CONSIMŢĂMÂNT INFORMAT

Vă rugăm să citiţi cu atenţie acest formular şi să întrebaţi coordonatorul studiului sau membrii echipei de cercetare a studiului orice informaţie pe care nu aţi înţeles-o clar. Vă rugăm să vă asiguraţi că toate întrebările dvs. au fost clarificate înainte să semnaţi

formularul. Informaţia obţinută din acest formular şi din discuţiile cu membrii echipei de cercetare vă poate ajuta să vă decideţi dacă doriţi să participaţi în acest studiu de cercetare. Dacă optaţi să participaţi, trebuie să vă scrieţi numele şi să semnaţi la sfârşitul formularului şi pe autorizarea de participare la cercetare.

INTRODUCERE

Studiul la care sunteţi invitat să participaţi urmăreşte să evalueze eficienţa intervenţiilor chirurgicale parodontale *versus* combinaţia între intervenţiile chirurgicale parodontale şi intervenţia psihologică cognitiv-comportamentală în tratamentul persoanelor cu boală parodontală sau retracţii gingivale izolate.

Este important să ştiţi că: (1) participarea la studiu este voluntară; (2) puteţi decide să nu luaţi parte sau puteţi să vă retrageţi din studiu în orice moment. În oricare caz, nu veţi pierde niciun beneficiu la care aveţi altfel dreptul. Însă, pentru a primi gratuit tratamentul prevăzut în studiu trebuie să acceptaţi participarea în studiu.

DESCRIEREA STUDIULUI

A. ELIGIBILITATE

Pot participa la studiu persoane cu vârste între 14-60 de ani, în bună stare de sănătate generală, care prezintă: ►una sau mai multe retracţii gingivale (RG) ≥ 2mm, clasa 1, 2 sau 3 Miller, fără intervenţii muco-gingivale anterioare în zona vizată sau ►parodontite cronice sau agresive, care necesită un tratament chirurgical de reducere a pungilor parodontale. Criteriile de excludere se referă la: persoanele cu afecţiunile sistemice care contraindică intervenţia chirurgicală, mobilitate dentară de gradul 3 sau 4, motivaţie insuficientă.

B. PROCEDURĂ

Bazat pe acordul dvs. voluntar şi verificarea prealabilă a eligibilităţii, puteţi participa în acest studiu care implică o fază de intervenţie chirurgicală şi o fază de urmărire a evoluţiei dvs. timp de 1 an, ambele gratuite. Toţi participanţii din studiu vor fi evaluaţi înaintea intervenţiei medicale, iar apoi la 14 zile, 1 lună, 3 luni, 6 luni şi 1 an.

Participanţii eligibili pentru studiu vor fi distribuiţi în grupul de intervenţie chirurgicală parodontală. Intervenţia chirurgicală parodontală constă în acoperirea zonelor de retracţie izolată prin procedeul cel mai indicat cazului sau printr-o terapie chirurgicală de reducere a pungilor, cu sau fără adiţie de os. Ulterior acestei intervenţii este urmărită evoluţia dvs. prin întâlniri stabilite la interval de 14 zile, 1 lună, 3 luni, 6 luni şi 1 an. Intervenţia menţionată anterior poate fi primită contra cost în majoritatea serviciilor medicale, atât publice, cât şi private.

Participanţii la studiu sunt rugaţi să completeze la înrolare şi pe parcursul derulării studiului diverse scale psihologice[9].

C. RISCURI

Efectele secundare ale intervenţiei chirurgicale parodontale se referă la efecte inerente asociate intervenţiei (dureri la locul intervenţiei, tumefieri) şi până la situaţii extreme, rar menţionate de literatura de specialitate (necroza grefei, hemoragii, expunerea şi infecţia membranei, atunci când este folosită). Tratamentul va fi oprit dacă apar unele dintre efectele mai sus menţionate.

D. BENEFICII

Beneficiul potenţial al implicării în acest studiu se referă în principal la o reducere a simptomelor asociate bolii de bază pentru care v-aţi adresat medicului, a simptomelor asociate intervenţiei. Deşi nu se poate spune că veţi beneficia în procent de 100% în urma acestor tratamente, studiile anterioare arată că un procent foarte mare de pacienţi îşi rezolvă sau îşi ameliorează semnificativ simptomele şi problemele asociate.

E. CONFIDENŢIALITATE

Acordul dvs. cu acest formular va autoriza echipa de cercetare să colecteze şi să folosească informaţii din studiu. Ţinând cont că informaţia colectată va fi utilizată numai pentru scopuri de cercetare, toate eforturile vor fi mobilizate pentru a nu vă dezvălui identitatea în orice document sau publicaţie rezultate în urma acestui studiu. Confidenţialitatea informaţiilor referitoare la dvs. va fi în acord cu legile aplicabile.

F. RETRAGEREA

Participarea este complet voluntară. Manifestarea interesului pentru studiu nu vă obligă să participaţi. Chiar dacă la început decideţi să participaţi şi apoi vă răzgândiţi, vă puteţi retrage în orice moment fără penalizări. Dacă veţi lua această decizie vă recomandăm: (1) să anunţaţi decizia dvs. (celui care v-a oferit tratamentul) pentru buna derulare a studiului; (2) să consultaţi obligatoriu un alt medic stomatolog pentru a stabili un alt tratament şi a preveni complicaţiile care ar putea să apară ţinând cont de condiţia dvs. De asemenea, pe parcursul cercetării cel care face tratamentul poate să vă recomande consultarea medicului de familie sau a unui alt medic stomatolog pentru a preveni anumite complicaţii; pentru reducerea riscurilor vă rugăm să luaţi în considerare aceste recomandări.

Echipa de cercetare are de asemenea dreptul de a opri participarea dvs. în studiu. Această decizie poate fi luată din cauza unor reacţii neaşteptate, a unor eşecuri repetate de a urma instrucţiunile sau din cauză că întreg studiul este oprit.

G. RESPONSABILITATE

Acest studiu a fost iniţiat de către Universitatea „Babeş-Bolyai” (Catedra de Psihologie Clinică şi Psihoterapie), Universitatea de Medicină şi Farmacie „Iuliu Haţieganu” (Catedra de Parodontologie), Universitatea de Medicină şi Farmacie „Victor Babeş” din Timişoara şi Institutul de Cercetări pentru Instrumentaţie Analitică Cluj. Calitatea şi responsabilitatea tratamentului revine medicului stomatolog care oferă direct tratamentul. Tratamentul va fi oferit în cadrul unui orar stabilit de comun acord cu dvs. Am citit explicaţiile privind acest studiu şi am avut oportunitatea de a adresa întrebări de clarificare adiţionale. Îmi dau acordul pentru participarea la studiu. Nume: ____________________________ Semnătură: ________________________ Dată: _____

Semnarea formularului de consimţire informată reprezintă ultima secvenţă a primei etape, de recrutare şi selecţie, specifică unui studiu randomizat controlat. Desigur, etapa de recrutare şi selecţie a participanţilor poate fi considerată drept etapă iniţială, având în vedere implementarea efectivă a studiului. Implementarea este precedată, de etapa de planificare a studiului, de etapa de obţinere a acordului pentru derularea studiului din partea unei comisii de etica cercetării şi, desigur, de etapa obţinerii fondurilor necesare pentru acoperirea cheltuielilor cu derularea studiului, considerate a fi etape premergătoare implementării.

4.2.1.2. Etapa de derulare a studiului

În cadrul acestei etape are loc alocarea randomizată a participanţilor în grupele din cadrul experimentului, se aplică intervenţiile convenite şi au loc măsurători intermediare (aspect opţional), respectiv cele realizate la finalul derulării intervenţiei (evaluare postintervenţie).

4.2.1.2.1. Alocarea randomizată în grupele experimentale

Secvenţa de alocare randomizată a participanţilor a fost descrisă în secţiunea 4.1.3. De aceea, ne vom rezuma la a prezenta cele trei caracteristici ale unei alocări randomizate în grupele experimentale considerate a fi necesare pentru o randomizare de o calitate superioară.

Primul aspect vizează algoritmul de randomizare ales. Acesta ar trebui să fie unul adaptat numărului de participanţi de inclus în fiecare grupă, tipului de recrutare (recrutaţi simultan sau secvenţial – vezi figura 4.3) şi să nu permită anticiparea grupei din care face parte un individ (de exemplu, nu este recomandabilă randomizarea sistematică, cu pas, deoarece ordinea de introducere a participanţilor în studiu ar putea fi manipulată de către cercetător, fapt care ar conduce la o distorsiune în selecţie).

Cel de-al doilea aspect dezirabil vizează separarea activităţii de stabilire a gradului de eligibilitate a unei persoane pentru a fi inclusă în studiu, de cea care presupune randomizarea participanţilor în grupele experimentale. Acest lucru se obţine prin alocarea acestor activităţi unor persoane diferite pentru a reduce riscul unei distorsiuni în selecţie (de exemplu, pentru a evita practica de a-i include pe cei eligibili, dar cu un profil care indică o complianţă scăzută la tratament, să fie incluşi în grupul de control pentru a maximiza şansele obţinerii unor diferenţe semnificative statistic, în direcţia dorită, aceea a superiorităţii lotului experimental).

Cea de-a treia caracteristică dezirabilă vizează transparenţa intervenţiilor (condiţia de orbire). Pentru a reduce riscurile apariţiei unor distorsiuni (efect Hawthorne, demoralizare, expectanţe diferite privind evoluţia participanţilor, comportament diferenţiat faţă de participanţi în funcţie de intervenţia primită, distorsiuni în evaluare) ar fi ideal dacă studiul ar fi derulat în condiţii de tip triplu-orb. Deoarece este puţin

probabil ca acest lucru să fie făcut în cercetări din domeniul psihologiei, mult mai uşoară ar fi implementarea unei condiţii de tip simplu-orb, având drept protagonişti asistenţii de cercetare responsabili pentru evaluările realizate preintervenţie, în timpul intervenţiei, postintervenţie sau în *follow-up*. Acest lucru ar fi necesar mai ales în contextul în care variabilele dependente principale sunt evaluate într-o manieră subiectivă, prin aprecieri ale progreselor făcute.

4.2.1.2.2. Implementarea intervenţiei

Această secvenţă trebuie să respecte aspectele planificate şi prezentate în formularul de consimţire informată (de exemplu, durata perioadei de intervenţie, gradientul intervenţiei – numărul de şedinţe). În plus, în această secvenţă există diferenţe importante între studiile cauzale (*efficacy studies*) şi cele pragmatice (*effectiveness studies*). În cazul primelor este foarte important să existe o respectare a protocolului iniţial atât din partea celor desemnaţi să implementeze intervenţia, în baza celor cuprinse într-un manual al intervenţiei, cât şi din partea celor care au consimţit să primească intervenţia alocată. De asemenea, conţinutul fiecărei întâlniri (şedinţe de implementare) trebuie respectat.

Din perspectiva implementării unei intervenţii, ar fi necesar să se observe măsura în care livrarea intervenţiei[10] a avut loc conform planului[11], respectiv măsura în care participanţii la studiu au aderat (au manifestat complianţă) la intervenţia primită. Aceste aspecte sunt necesare pentru a aprecia dacă sunt afectate validitatea de construct (livrarea intervenţiei), respectiv validitatea internă a studiului (aderenţa la intervenţie). În cazurile în care se constată abateri substanţiale în implementarea intervenţiei de la cele prevăzute iniţial, se poate ajunge la propuneri de excludere a acestor cazuri din analiză, deoarece devierea

semnificativă de la planul iniţial constituie un factor de risc pentru validitatea internă şi validitatea de construct a concluziilor studiului. De aceea, acest control strict prezent în studiile cauzele este necesar pentru a reduce erorile sistematice, în ideea de a ne asigura că intervenţia (aşa cum a fost ea planificată) este cea care a condus la efectele obţinute şi nu alte aspecte colaterale, introduse pe parcurs.

În schimb, studiile pragmatice pornesc de la premisa că este mai importantă reproducerea condiţiilor din realitate (în care trainerii, psihoterapeuţii, deşi au un plan iniţial, pot devia de la acesta în funcţie de nevoile fiecărui participant), decât să se respecte aspectele planificate, deşi acest lucru creşte riscul de a oferi intervenţii ce variază, între anumite limite, de la un participant la altul, cu consecinţe negative asupra validităţii interne şi de construct a concluziilor studiului. În schimb, prin această strategie adoptată, studiile pragmatice câştigă în ceea ce priveşte validitatea externă a studiului, îndeosebi cu privire la validitatea ecologică.

De altfel, pe lângă diferenţele ce ţin de condiţiile de eligibilitate pentru includerea în studiu (mai relaxate în cazul studiilor pragmatice), principala diferenţă dintre studiile cauzale şi cele pragmatice are în vedere modul de implementare a intervenţiilor. Cerinţa ca implementarea intervenţiei să se realizeze conform celor explicitate în manualul intervenţiei este mult mai strictă în cazul studiilor cauzale decât în a celor pragmatice.

Pentru a creşte fidelitatea faţă de propunerea iniţială, cu ocazia implementării unei intervenţii se pot utiliza o serie de strategii menite să sprijine validitatea de construct (Shadish *et al.*, 2002): (a) în direcţia livrării corecte a intervenţiei, prin existenţa unui manual de implementare a intervenţiei, susţinut de mijloace instructive (casete video

demonstrative, discuţii şi feedback corectiv din partea unor supervizori, în baza unor cazuri înregistrate; training în implementare); (b) în direcţia creşterii aderenţei (complianţei) participantului la intervenţie/tratament, prin comunicarea explicită a temelor pentru acasă însoţită de hand-out-uri, quizz-uri periodice de cunoştinţe, discuţii/discursuri motivaţionale, discuţii cu membrii familiei în ideea încurajării implicării în intervenţie, stimularea aderenţei la intervenţie prin oferirea de mici premii (tombolă) etc.

Un aspect extrem de important, dar deseori neglijat în partea de implementare, se referă la identificarea influenţei unor factori externi deoarece validitatea concluziilor unui studiu controlat nu ţine doar de implementarea corectă a intervenţiei, ci şi de absenţa unor factori străini perturbatori. Astfel, este posibil ca participanţii să fie incluşi concomitent în mai multe studii şi să existe, astfel, o interferenţă sau un transfer între intervenţiile primite. De asemenea, trebuie verificată situaţia celor incluşi în grupul de control pasiv (fără intervenţie), pentru a observa măsura în care aceştia îşi menţin acest statut pe parcursul intervenţiei sau dacă acţionează individual în vederea ameliorării problemei lor[12].

Pe lângă importanţa livrării acurate a intervenţiei un alt factor important, trebuie luat în calcul în legătură cu cei care implementează intervenţia, vizează competenţa acestora de a oferi o intervenţie de calitate. În acest sens, se obişnuieşte să se apeleze la persoane care: (a) deţin certificare recunoscută în domeniul intervenţiei (de exemplu, certificat de formator în trainingul sau coaching-ul respectiv, atestat de psihoterapeut în tehnica psihoterapeutică implementată); (b) posedă o experienţă relativ bogată, preferabil atât în aplicarea intervenţiei respective, cât şi în abordarea problemelor supuse spre rezolvare (de exemplu, experienţă în cazuistică depresivă). Desigur, acest aspect nu are

un caracter obligatoriu atât timp cât este monitorizată atent partea de livrare a intervenţiei, mai ales dacă se au în vedere aspecte economice (raportul cost-eficienţă). Se cunoaşte faptul că nivelul costurilor unei intervenţii creşte dacă pentru implementarea acesteia se apelează la specialişti decât dacă ea se bazează pe novici.

În sfârşit, pe parcursul derulării intervenţiei ar fi indicat să se ia măsuri care să conducă la un procent cât mai redus de renunţare la cercetare (*dropout* sau *attrition*). De exemplu, ar fi dezirabil, în cazul în care există mai mulţi asistenţi de cercetare, să fie alocată aceeaşi persoană pentru a colecta datele cu ocazia evaluărilor diverse (pentru a ajuta la consolidarea relaţiei interpersonale dintre participanţi şi asistenţi, desigur, dacă nu au existat conflicte între cele două părţi). De asemenea, i se poate aduce aminte periodic participantului despre importanţa studiului pentru cunoaştere sau i se pot oferi stimulente pentru participare (inclusiv mici cheltuieli de protocol – apă, fursecuri). În plus, pentru a preveni absenteismul de la şedinţele de intervenţie, se poate trimite un SMS sau e-mail de reamintire cu privire la data, ora şi locaţia următoarei şedinţe. O altă măsură deosebit de utilă este aceea de a reduce perioada de timp existentă între semnarea formularului de consimţire informată şi implementarea efectivă a intervenţiei.

4.2.1.2.3. Evaluări colectate cu ocazia derulării intervenţiei

Orice studiu randomizat controlat cuprinde obligatoriu o secvenţă de evaluare realizată postintervenţie, pentru a urmări diferenţele observate între grupele incluse în studiu. Această evaluare este necesară[13] pentru a demonstra existenţa unei relaţii cauzale. În funcţie de natura grupului de control, rezultatul obţinut la nivel de postintervenţie poate fi utilizat pentru a arăta că: (a) intervenţia este una eficientă (control exercitat prin

liste de aşteptare/absenţa intervenţiei); (b) efectul se datorează caracteristicilor specifice ale intervenţiei (control exercitat prin placebo/intervenţii nespecifice); (c) intervenţia este mai bună, cel puţin la fel de bună sau nu este inferioară alteia (control exercitat prin intervenţii uzuale, intervenţii recomandate, intervenţii standard) etc. Desigur, pentru a potenţa comparaţiile făcute şi efectele obţinute, modelul standard de studiu randomizat controlat presupune evaluări ale variabilei dependente primare, atât la nivel de preintervenţie, cât şi la finalul perioadei de urmărire, adică la un anumit interval de timp după încheierea aplicării intervenţiei.

În schimb, dacă obiectivul urmărit prin studiu nu este doar acela de a demonstra existenţa unei relaţii de tip cauză-efect, ci vizează identificarea unor mecanisme psihologice responsabile pentru schimbarea obţinută, atunci etapa de derulare a intervenţiei va cuprinde suplimentar, pe lângă evaluarea VD, măsurători şi ale potenţialelor variabile mediatoare (cu rol explicativ pentru schimbările observate). În plus, pe lângă momentele standard (preintervenţie, postintervenţie şi finalul perioadei de urmărire), se apelează la una sau mai multe evaluări intermediare (pe parcursul intervenţiei), atât privind nivelul variabilelor dependente, cât şi cel al potenţialelor variabile mediatoare (mecanisme ale schimbării).

Tabelul 4.5. *Momente ale evaluării în funcţie de obiectivul studiului*

	Mecanismele schimbării Evaluarea mediatorilor				Evaluarea relaţiei cauză-efect Evaluarea VD			
Tip de design	Pre	Interim	Post	Urm	Pre	Interim	Post	Urm
Standard (demonstrarea unei relaţii cauzale)	O	O	O	O	X	O	X	X
Identificarea mecanismelor schimbării	X	X	X	X	X	X	X	X

Legendă: X – evaluare; O – absenţa evaluării

Aceste evaluări intermediare reprezintă una dintre condiţiile necesare pentru a identifica un anumit mecanism al schimbării („cum se produce schimbarea, efectul?”), deoarece premisa acceptată este că modificările la nivelul variabilei mediatoare precedă temporal modificările de la nivelul variabilei dependente primare. De pildă, într-un studiu ce urmăreşte eficienţa terapiei raţional-emotive în combaterea depresiei, în care se teoretizează că principalul mecanism al schimbării pentru reducerea simptomatologiei depresive vizează diminuarea cogniţiilor iraţionale autodepreciative, ar trebui introdusă cel puţin o evaluare intermediară, însă dezirabil ar fi să existe mai multe asemenea evaluări intermediare. Aceasta ne-ar permite să observăm că, pe măsura implementării terapiei, apare mai întâi o diminuare a nivelului cogniţiilor iraţionale autodepreciative urmată de o reducere a simptomatologiei depresive şi că între cele două aspecte există o asociere semnificativă. Astfel, în studiile focalizate pe mecanismele schimbării se apelează la evaluări intermediare (pe parcursul derulării intervenţiei), nu numai ale variabilelor dependente de interes, ci şi ale aspectelor psihologice care ar putea fi identificate drept mecanisme ale schimbării (care explică de ce este eficientă intervenţia şi cum poate fi optimizată aceasta). Alte detalii despre condiţiile necesare pentru a identifica mecanismele schimbării (variabilele mediatoare) responsabile pentru producerea efectului sunt prezentate în secţiunea 4.2.3.

Un alt aspect legat de evaluarea VD se referă la cutuma de a diferenţia, în cadrul studiilor randomizate controlate, între variabilele dependente principale (*primary outcomes*) şi variabilele dependente secundare. De exemplu, într-un studiu care are drept scop implementarea unei intervenţii care să reducă nivelul de epuizare emoţională (*burnout*)

printre angajaţi, nivelul de *burnout* reprezintă variabila dependentă primară. Desigur, cu ocazia acestui studiu pot fi colectate şi alte date, cum ar fi nivelul productivităţii în muncă, satisfacţia legată de locul de muncă, implicarea în muncă etc.), însă acestea reprezintă aspecte colaterale care pot fi afectate de intervenţia implementată. Variabila dependentă primară are rolul de a oferi răspunsul principal cu privire la testarea ipotezei („este intervenţia una eficientă pentru problema pentru care a fost dezvoltată?”), precum şi în luarea unor decizii cu impact metodologic cum ar fi stabilirea numărului de participanţi în studiu, în baza efectului estimat al intervenţiei asupra variabilei dependente principale.

În sfârşit, un ultim aspect de discutat, cu relevanţă pentru evaluarea impactului intervenţiei, are în vedere diferenţele existente între studiile cauzale şi cele pragmatice cu privire la alegerea variabilelor dependente măsurate. Pentru a prezenta aceste diferenţe, vom distinge mai întâi între variabile dependente reale (*true outcomes*) şi variabile dependente surogat (*surrogate outcomes*) (Torgerson şi Torgerson, 2008). De pildă, în cazul unui program menit să reducă delincvenţa la adolescenţi, variabile dependente reale ar putea fi prezenţa sau absenţa unui comportament delincvent înregistrat la poliţie, prezenţa sau absenţa unei recidive, în timp ce variabile dependente surogat ar putea fi nivelul de agresivitate sau nivelul comportamentului antisocial măsurat prin scale de specialitate. VD reale sunt cele vizate direct sau indirect de orice intervenţie, prin impactul lor la nivel social. În schimb, VD surogat, deşi se află în strânsă legătură cu aspectele reale vizate şi deseori conduc la apariţia VD reale, nu oferă certitudinea apariţiei acestor efecte. Spre exemplu, nu întotdeauna o satisfacţie maritală scăzută se soldează cu divorţ, nu întotdeauna eşecul de a reduce cogniţiile iraţionale asociate violenţei

domestice se manifestă prin acte manifeste de violenţă domestică, nu întotdeauna reducerea densităţii oaselor se manifestă prin fracturi etc.

Tabelul 4.6. *Exemple de variabile dependente reale, respectiv surogat în domenii diverse*

Scopul intervenţiei	VD surogat	VD reală
Probleme în relaţia de cuplu	satisfacţie maritală	divorţ
Violenţa domestică	reducere cogniţii iraţionale	reducerea agresivităţilor raportate
Osteoporoza	reducerea densităţii din oase	fractură
Domeniul educaţional	atitudinea faţă de şcoală	abandon şcolar
Domeniul organizaţional	epuizare emoţională	procent fluctuaţie de personal
Infracţionalitatea	nivelul de ostilitate/agresivitate	comitere infracţiuni

Studiile pragmatice, prin caracteristica lor de a avea o validitate externă (ecologică) ridicată, prezintă un interes crescut pentru publicul larg şi pentru formatorii de politici. De aceea, în cazul lor se recomandă să se utilizeze variabile dependente reale. Acest lucru este posibil şi pentru că, în cazul acestui gen de studii (condiţii relaxate de eligibilitate/absenţa monitorizării integrităţii intervenţiei), pot fi colectate mai multe date (mai mulţi participanţi), care să asigure o putere statistică suficientă pentru a detecta mărimi ale efectelor mai scăzute, mai specifice cazului măsurării VD reale decât măsurării variabilelor-surogat.

În schimb, în studiile cauzale, deşi ar fi dezirabil să fie utilizate tot VD reale în calitate de VD principale, acest lucru deseori nu este posibil din raţiuni metodologice. Prin condiţiile stricte impuse cu privire la selectarea participanţilor şi implementarea studiului este dificil să fie strâns un volum foarte mare de participanţi la studiu. În plus, efectele oricărei intervenţii tind să fie mai ridicate ca mărime în cazul variabilelor-surogat decât a celor reale. În consecinţă, pentru a evita probleme cauzate de o putere statistică mai scăzută (de genul erorii de tip II în testarea

ipotezelor), se obişnuieşte ca VD principale într-un studiu cauzal să fie din categoria VD surogat, chiar dacă sunt colectate secundar şi măsuri ale VD reale.

4.2.1.3. Etapa de urmărire

Scopul implicit al oricărei intervenţii este acela de a produce modificări în direcţia dorită care să reziste în timp, astfel încât efectele produse să nu fie doar unele tranzitorii. De aceea, în sprijinul acestui scop nu este suficient doar să ne focalizăm asupra comparaţiei dintre grupe, efectuată cu ocazia evaluării postintervenţie (de la finalul etapei de derulare a intervenţiei), ci este nevoie să urmărim (*follow-up*) dacă diferenţele observate se menţin în timp, se accentuează sau se reduc sau chiar dispar, odată cu trecerea timpului.

Durata etapei de urmărire nu este una standard, însă majoritatea studiilor randomizate controlate apelează la o perioadă cuprinsă între şase luni şi doi ani. Rolul cercetătorului în această secvenţă este în principal acela de a apela la măsuri care să asigure un nivel cât mai scăzut de pierdere a participanţilor, îndeosebi prin: (a) asigurarea accesului/disponibilităţii participanţilor pentru evaluări în perioada de urmărire; (b) verificarea măsurii în care în istoricul participantului nu au fost înregistrate evenimente care să afecteze validitatea concluziilor obţinute (de exemplu, verificarea faptului că participantul nu a apelat în perioada de urmărire la intervenţii/tratamente alternative pentru problematica vizată, nu a apelat la alţi specialişti pentru a continua să primească intervenţia alocată etc.).

Pentru primul aspect, acela de a facilita accesul la participanţi în perioada de *follow-up*, câteva indicaţii utile (Shadish *et al.*, 2002) ar fi: (a)

colectarea datelor de contact nu numai de la participant, ci şi de la una sau două persoane apropiate acestuia, pentru a creşte şansa de a-l contacta ulterior, în cazul în care datele sale de contact nu mai sunt valabile; (b) menţinerea contactului cu participanţii prin scurte informări telefonice (scrise), pentru a obţine reconfirmarea de la participant cu privire la disponibilitatea sa de a fi evaluat ulterior, îndeosebi în perioada premergătoare evaluării (de exemplu, cu o săptămână înainte); în plus, preferabil ar fi dacă comunicarea ar fi făcută de către acelaşi asistent de cercetare cu care participantul a interacţionat pe parcursul evaluărilor anterioare; (c) flexibilitate cu privire la locaţia în care are loc evaluarea (deplasare la sediul participantului) şi/sau oferirea de stimulente (plata transportului, recompensare financiară pentru şedinţa de evaluare); (d) mici gesturi de atenţie (de exemplu, trimiterea unei felicitări cu ocazia zilei de naştere a participantului); (e) oferirea unei broşuri cu informaţii utile (calendarul evaluărilor în perioada de *follow-up*; menţionarea stimulentelor financiare oferite, dacă este cazul, pentru evaluările din perioada de urmărire; precizarea datelor de contact unde participantul ar trebui să anunţe dacă apar modificări în datele sale de contact etc.).

În ceea ce priveşte cel de-al doilea aspect care conduce la pierderea participanţilor, acela de contaminare a intervenţiei alocate (de exemplu, persoane care au urmat intervenţia X în etapa de derulare a acesteia, dar care apelează la intervenţia alternativă Y în perioada de urmărire), se pot activa câteva strategii de reducere a acestui risc. Una dintre acestea constă în oferirea posibilităţii participantului de a apela la câteva (două sau trei) şedinţe suplimentare din intervenţia alocată (*boosting sessions*). De exemplu, în cazul intervenţiilor din zona psihoterapiei este acceptabil să li se ofere două-trei şedinţe suplimentare într-un interval de şase luni, pe care participanţii să le planifice în perioada de urmărire, doar dacă au

neapărat nevoie de ele. O altă posibilitate existentă, fezabilă în anumite cazuri (în funcţie de problema şi aspectele etice implicate), este aceea de a include participantul pe o listă de aşteptare şi de a-i oferi ulterior intervenţia promisă, după finalizarea perioadei de urmărire.

Desigur, o măsură obligatorie în legătură cu problema contaminării intervenţiei, este aceea de a intervieva participanţii (de exemplu, telefonic, cu ocazia pregătirii şedinţelor de evaluare derulate în perioada de urmărire) pentru a vedea în ce măsură ei au apelat pe cont propriu la continuarea intervenţiei respective sau la alte metode de intervenţie concurente (alternative). Asemenea situaţii îl scot pe participant din faza de urmărire deoarece primeşte în continuare un tratament activ. Tot cu ocazia acestui interviu pot fi evidenţiate evenimente importante apărute în istoricul individului, care ar putea afecta evoluţia sa (prin diminuarea sau accentuarea problemei avute).

4.2.1.4. Etapa de analiză

Din punct de vedere logistic, aceasta reprezintă cea mai uşoară şi scurtă etapă din cadrul unui studiu randomizat controlat. Din raţiuni pedagogice, vom diviza această etapă în trei elemente distincte: (a) analiza standard; (b) analiza la nivelului intenţiei de a interveni (*intent to treat*); (c) alte analize relevante.

4.2.1.4.1. Analiza standard

În cadrul acestei analize se testează ipotezele cercetării, un rol principal revenindu-i ipotezei care vizează efectul intervenţiei asupra VD primare. Din acest punct de vedere, cele mai importante analize sunt comparaţiile dintre grupe realizate la momentul postintervenţiei, precum şi în perioada

de *follow-up*. Comparaţia de la nivelul postintervenţiei ne arată dacă intervenţia propusă a avut un efect (este mai eficientă decât o intervenţie alternativă etc.), în timp ce comparaţiile la nivel de *follow-up* ne indică dacă efectul este unul durabil. În cazul analizei standard sunt relevante (vezi nota 1, p. 174) şi analizele intragrupale (de exemplu, pentru a analiza dacă există modificări semnificative ale VD între pretest şi posttest în cadrul grupului experimental).

Tot în cadrul analizei standard are loc şi analiza de verificare a echivalenţei grupelor incluse în studiu la nivel de pretest (preintervenţie). Dacă echivalenţa asumată în baza alocării randomizate este susţinută (dublată) empiric prin compararea evaluărilor efectuate la momentul preintervenţiei, datele de la nivel de postintervenţie şi *follow-up* pot fi analizate simplu, prin apelul la testul t student sau tehnicile ANOVA, în funcţie de caz. Dacă echivalenţa la nivel de preintervenţie nu este sprijinită empiric (există diferenţe între grupele comparate, în ciuda alocării randomizate), aceste diferenţe pot fi tratate drept covariabile sau pot fi incluse în analizele statistice standard ale diferenţelor observate între pretest şi posttest (*gain scores*).

Desigur aceste aspecte sunt valabile atât pentru studiile cauzale, cât şi pentru cele pragmatice, în contextul în care obiectivul studiului se rezumă la identificarea unei relaţii de tip cauză-efect. Dacă obiectivele studiului vizează identificarea mecanismelor schimbării, designul şi analiza standard se complică (vezi detalii în secţiunea 4.2.3).

4.2.1.4.2. Analiza la nivelul intenţiei de a interveni

Spre deosebire de cazul ideal, în realitate între numărul participanţilor imediat după momentul alocării lor randomizate în grupele experimentale şi numărul participanţilor la finalul perioadei de urmărire pot exista

diferenţe mari. Acest lucru se întâmplă pentru că unii participanţi refuză să urmeze intervenţia alocată, deşi iniţial şi-au dat acordul pentru aceasta, alţii renunţă pe parcursul derulării intervenţiei (nu finalizează programul convenit), iar alţii sunt pierduţi pe parcursul duratei de *follow-up* fie din cauza apariţiei contaminării intervenţiei, fie din cauza absenţei măsurătorilor din perioada de urmărire (participanţi care refuză să fie evaluaţi repetat sau participanţi ale căror date de contact nu mai sunt valabile şi care nu mai pot fi reperaţi prin mijloace alternative de căutare). În urma acestei situaţii este posibil ca uneori doar 50-60% dintre participanţii alocaţi iniţial într-un anumit grup experimental să poată fi evaluaţi până la final pentru a surprinde efectele intervenţiei.

Din acest punct de vedere, este important să se identifice motivul pentru care participantul a fost pierdut pentru analiză: (a) din cauza intervenţiei în sine (*treatment attrition*) (participantul renunţă pentru că ar fi preferat o altă intervenţie, nu o găseşte suficient de interesantă pentru a continua pe cea alocată, resimte efecte secundare negative etc.); (b) din motive obiective ale participantului (se mută într-o altă localitate, apariţia unor evenimente cu impact în viaţa sa); (c) din motive obiective tehnice (diminuarea finanţării, însoţită de o reducere a numărului de participanţi; defecţiuni în salvarea unor date pe calculator etc.). Dacă persoanele incluse în punctele b şi c pot fi excluse din analiză pe anumite secvenţe (acolo unde lipsesc datele) fără a afecta validitatea concluziilor, excluderea participanţilor care au renunţat la studiu din motivele invocate la punctul a poate conduce la rezultate distorsionate. Să ne imaginăm că am fi dorit să evaluăm în ce măsură studenţilor le-ar plăcea să studieze metodologia cercetării printr-o strategie didactică alternativă. Să presupunem că am fi inclus iniţial 100 de persoane, dar că pe parcurs 25 dintre ele au renunţat pentru că se plictiseau sau nu înţelegeau prelegerile,

iar alte 25 au renunţat în favoarea unei alte metode de predare pe care o găseau mai atractivă. Dacă noi am analiza, în acest caz, datele colectate doar de la cei 50 de participanţi rămaşi în sală până la final cu siguranţă am obţine scoruri distorsionate, mult mai favorabile intervenţiei propuse decât dacă am fi colectat datele de la toţi cei 100 de participanţi iniţiali.

Pentru a preîntâmpina asemenea distorsiuni, în cazurile în care se înregistrează un procent mai mare de 10-15% de renunţare din cauza incompatibilităţii cu intervenţia sau în care se constată diferenţe în rata de renunţare între grupul experimental (de exemplu, 15%) şi cel de control (de exemplu, 5%), se apelează complementar la analiza statistică la nivelul intenţiei de a interveni în defavoarea analizei standard, în plus faţă de analiza standard. Astfel, vom putea compara impactul pe care îl are fenomenul de pierdere a participanţilor cauzat din alte motive decât cele obiective şi în ce măsură afectează el concluziile studiului.

Soluţia uzuală în analiza de acest tip presupune luarea în considerare a ultimelor scoruri disponibile, inclusiv de la participanţii care au renunţat să primească intervenţia din dorinţa lor. De exemplu, dacă există date înregistrate doar la nivel de preintervenţie, se va asuma că aceste valori vor fi obţinute şi în testările ulterioare. Dacă participantul a fost evaluat iniţial (pretest), dar a participat la o parte dintre şedinţe, fiind evaluat intermediar în două momente (de exemplu, după şedinţa 4 şi după şedinţa 8), însă nu există rezultate cu privire la performanţa lui măsurată la momentul postintervenţiei, se va trece în locul datelor lipsă scorul înregistrat la ultima evaluare intermediară pentru care există date (şedinţa 8) etc.

O strategie alternativă mai bună, însă cu sorţi mai mici de izbândă, ar fi aceea de a solicita, inclusiv celor care au renunţat la intervenţia alocată din motive de incompatibilitate cu aceasta, să participe la evaluările

periodice (fie ele intermediare, postintervenţie sau din perioada de urmărire). Aceste rezultate ar reprezenta mai bine situaţia reală decât asumarea păstrării constante a rezultatelor obţinute cu ocazia ultimei evaluări disponibile. Desigur, cel mai probabil, ar fi nevoie de oferirea de stimulente pentru a-i determina pe cei care au renunţat la studiu din motive de incompatibilitate cu intervenţia alocată, să rămână participanţi activi în studiu, pentru această secvenţă de evaluare.

În sfârşit, un alt aspect important de urmărit, indiferent de procentul participanţilor care s-au retras din studiu, este analiza caracteristicilor participanţilor care s-au retras din studiu. De interes sunt atât caracteristicile lor demografice, cât mai ales unele variabile de interes (factori de risc, nivelul VD principale la pretest etc.), pentru a vedea care segment de participanţi este mai vulnerabil de a renunţa la tipul de intervenţie alocată.

4.2.1.4.3. Alte analize relevante

Cel puţin în domeniul psihologiei clinice, unde nivelul de rigurozitate asociat studiilor randomizate controlate este unul extrem de ridicat ca urmare a modelului preluat din ştiinţele medicale, se recomandă analize suplimentare celor care vizează direct evaluarea variabilelor dependente, a celor mediatoare sau moderatoare. În această direcţie se recomandă a se diferenţia între efectul/impactul factorilor comuni, nespecifici formei de intervenţie analizată (cum ar fi alianţa terapeutică, gradul de aderenţă la protocolul intervenţiei, complianţa clienţilor), de impactul factorilor specifici intervenţiei respective.

De exemplu, în cazul aderării la protocolul intervenţiei, se observă măsura în care activităţile sunt conforme cu cele precizate în protocolul

intervenţiei în cadrul unui eşantion de şedinţe din cele efectuate, pentru a oferi siguranţă suplimentară pentru validitatea de construct a rezultatelor. De asemenea, se poate diferenţia efectul specific al intervenţiei după controlarea variaţiilor în nivelul VD determinate de factorii psihoterapeutici nespecifici (relaţionarea şi încrederea în psihoterapeut, implementarea intervenţiei în acord cu recomandările pentru aceasta, complianţa clienţilor la tratament etc.). În această direcţie au fost dezvoltate o serie de scale care să măsoare aceste aspecte nespecifice, ce pot fi evaluate şi supuse analizei statistice în mod suplimentar faţă de aspectele directe urmărite.

Un alt aspect relevant este conceptul de semnificaţie clinică (care ar putea fi extins către alte domenii de interes sub numele de semnificaţie sau relevanţă practică). La ce se referă acest lucru? Obţinerea unei diferenţe semnificative statistic între două grupe comparate (de exemplu, între un grup care a primit o intervenţie menită să reducă nivelul agresivităţii şi un grup de control de tipul listei de aşteptare) nu înseamnă automat că intervenţia este deosebit de benefică din punct de vedere practic. E posibil, de exemplu, ca nivelul de agresivitate în grupul experimental să fi scăzut ca frecvenţă de apariţie, dar din punct de vedere practic nu prea are relevanţă dacă cineva apelează la un limbaj agresiv cu o frecvenţă de 7 ori pe oră sau cu o frecvenţă de 5 ori pe oră. În mod similar, dacă nivelul de depresie în urma unei intervenţii se modifică de la 17 la 15 puncte (măsurate cu ajutorul scalei Beck), individul suferă în continuare de un nivel moderat de depresie, chiar dacă diferenţa de două puncte replicată la nivelul celor două grupe ar însemna o diferenţă semnificativă statistic. De aceea, pe lângă observarea unei diferenţe între grupele comparate reflectate printr-un rezultat semnificativ statistic ($p < 0,05$), o relevanţă anume are şi intensitatea acestei modificări – cunoscută

în lumea statistică sub denumirea de mărimea efectului. Semnificaţia clinică nu este însă dată automat de către nivelul mărimii efectului, deşi în cele mai multe cazuri există o corespondenţă între cele două. Pentru a înţelege acest lucru, să presupunem că o intervenţie pentru slăbit are o intensitate ridicată, ca mărime a efectului, în medie nivelul greutăţii corporale scăzând cu 15 kg. Din punct de vedere al semnificaţiei clinice, această diferenţă are o însemnătate diferită dacă intervenţia este aplicată persoanelor uşor supraponderale (cum ar fi o scădere în greutate de la o medie de 75 de kg la o medie de 60 de kg, decât dacă ar fi întâlnită în cazul persoanelor cu un nivel ridicat de obezitate (un lot de femei care ajung la 100 de kg de la 115 kg). Deşi magnitudinea schimbării este aceeaşi în ambele situaţii, din punct de vedere al semnificaţiei clinice, intervenţia este mai puţin valoroasă în cel de-al doilea caz discutat, deoarece calitatea vieţii acestor persoane nu se schimbă mult. Aşadar semnificaţia clinică are în vedere modificări importante în starea celui evaluat. Acestea ar putea fi diferenţiate în trei categorii distincte: clienţi cu o situaţie înrăutăţită sau neschimbată; clienţi cu o situaţie ameliorată (încă sunt probleme, dar reduse); respectiv pacienţi recuperaţi (problemele nu mai prezintă o intensitate clinică, VD măsurată fiind în zona de normalitate).

Pentru mai multe detalii despre acest subiect, inclusiv despre modalitatea efectivă de calcul a unor asemenea indicatori ai semnificaţiei clinice, cum este metoda mixtă Jacobson-Truax, ce combină comparaţiile normative (1) sau (2) cu nivelul magnitudinii schimbării (3) (vezi tabelul 4.7), poate fi consultată lucrarea lui Pintea (2010).

Tabelul 4.7. *Exemple de modalităţi de stabilire a semnificaţiei clinice*

Scopul intervenţiei	**Exemplificare**	**Avantaje şi dezavantaje**

Efectuarea de comparaţii normative Plasarea în baza unui etalon (1)	Considerarea cazurilor ca fiind recuperate dacă se încadrează sub percentilul 90 al unei probleme, respectiv ameliorate, dacă scorul lor la postintervenţie este între percentilele 90 şi 95.	Nu necesită alte măsurători înafara VD. Poate fi implementată doar în cazul măsurătorilor pentru care există etaloane adecvate.
Efectuarea de comparaţii normative Raportarea la criteriile de diagnostic (2)	Considerarea cazurilor ca fiind recuperate dacă nu mai îndeplinesc criteriile pentru tulburarea/problema respectivă.	Nu necesită decât măsurători ale VD. Supraestimează efectul pentru cei cu probleme uşoare, respectiv subestimează pentru indivizii cu probleme severe. Aplicabilitate doar în anumite cazuri de VD.
Aprecierea gradului de îmbunătăţire Indicatori ai magnitudinii schimbării (3)	Indexul modificării reale (*reliable change index*) vizează măsura în care ne putem încrede că modificarea observată nu putea fi produsă de variaţii aleatorii, ci din cauza intervenţiei.	Nu necesită alte măsurători, decât VD. Poate fi implementată în orice context. Supraestimează efectul intervenţiei pe loturi omogene (cu amplitudine redusă a VD la nivel de pretest), respectiv subestimează efectul pentru loturi eterogene.
Evaluări subiective Indicatori ai calităţii vieţii/ai satisfacţiei (4)	Utilizarea unor scale care să indice nivelul de funcţionalitate în diverse arii: social, profesional etc., inclusiv a unor standarde (*benchmarks*) pentru ceea ce ar presupune funcţionare normală.	Uşor de implementat, însă rezultatul este dependent şi de alţi factori care pot contribui la eşecul/succesul observat. Identifică percepţia individului şi nu situaţia reală (existând diferenţe de percepţie a modificărilor în funcţie de nivelul aşteptărilor clienţilor).
Impact social Indicatori sociali (5)	Monitorizarea unor indicatori relevanţi într-o anumită perioadă de timp: încarcerare (da/nu); recidivă (da/nu); abandon şcolar (da/nu); studii universitare (da/nu); status angajare (angajat/nu) etc.	Sunt indicatorii cei mai apreciaţi/relevanţi de către cei care dezvoltă politici publice. Necesită o perioadă lungă de timp pentru a observa impactul. Rezultatul este dependent şi de alţi factori.

4.2.2. Specificul studiilor randomizate la nivel de cluster

În domeniul clinic este destul de uşor să se implementeze studii randomizate la nivel individual deoarece adesea intervenţia are loc la nivel (a) individual şi (b) în afara contextului natural al individului (de obicei într-o clinică sau un cabinet de psihoterapie). În alte domenii din sfera psihologiei, precum cel educaţional şi cel organizaţional, intervenţiile nu pot fi aplicate în asemenea condiţii din motive variate. În schimb, în asemenea situaţii, intervenţiile pot fi implementate mai uşor la nivel (a) de grup şi (b) în contexte naturale (la şcoală sau la locul de muncă). Aceste caracteristici fac dificilă, dacă nu imposibilă, alocarea randomizată a indivizilor (la nivel individual) în grupele experimentale. De aceea, în multe cercetări efectuate în domeniul educaţional, respectiv cel organizaţional, intervenţiile au fost evaluate prin studii de tip cvasiexperimental (vezi secţiunea 4.2.4 pentru detalii), în care comparaţia dintre grupul experimental şi cel de control s-a făcut fără asigurarea evitării distorsiunii de selecţie, aplicând intervenţia pe grupurile naturale existente (cum ar fi clasa a IX-a A *vs* clasa a IX-a B; departamentul X *vs* departamentul Y), cu consecinţe negative asupra validităţii interne a acestor studii (a încrederii în concluziile obţinute). Aşa cum s-a putut deduce, atât din piramida încrederii în rezultatele studiilor prezentată în capitolul al doilea, cât şi din aspectele despre validitate discutate în cel de-al treilea capitol, studiile de tip cvasiexperimental sunt inferioare celor de tip experimental.

Soluţia pentru depăşirea acestui impas determinat de condiţiile obiective existente în contexte educaţionale sau organizaţionale, valabilă şi pentru alte contexte în care intervenţiile se realizează la nivel de grup,

este aceea de a apela în locul studiilor cvasiexperimentale la studii experimentale în care alocarea randomizată a participanţilor nu se face la nivel de individ, ci la nivel de grup/echipă/şcoală etc., denumit generic cluster.

Dincolo de aceste situaţii obiective care conduc la utilizarea randomizării la nivel de cluster, merită precizată situaţia în care lucrul la nivel de cluster se poate dovedi mai potrivit decât apelul la randomizare individuală. D.J. Torgerson şi C.J. Torgerson (2008) consideră mai potrivită randomizarea la nivel de cluster decât cea la nivel de individ pentru a reduce riscul contaminării (de exemplu, situaţii în care participanţii din grupul experimental pot povesti celor din grupul de control şi îi pot influenţa/transfera abilităţi dobândite în cadrul intervenţiei sau în cazul în care cei din grupul de control se simt demoralizaţi că nu beneficiază de aceleaşi „facilităţi" ca cei din grupul experimental).

De asemenea, spre deosebire de randomizarea la nivel individual, în alocarea randomizată a participanţilor la nivel de cluster există mai multă flexibilitate în modul de realizare a evaluărilor periodice efectuate cu scopul de a urmări efectul produs de intervenţie. În randomizarea la nivel de individ este necesar ca fiecare individ să fie testat ori de câte ori sunt prevăzute asemenea evaluări pe parcursului studiului (cum ar fi înainte de intervenţie, postintervenţie sau după finalizarea perioadei de urmărire). În contextul unei alocări randomizate a grupurilor (clusterelor) nu este obligatoriu nici să fie testaţi toţi indivizii din grupul respectiv, nici să fie evaluaţi aceiaşi indivizi selectaţi (cu riscul de a se obişnui cu testarea respectivă, fapt ce reduce acest tip de risc la nivelul validităţii interne a studiului). Astfel, în studii experimentale cu alocarea randomizată la nivel de cluster pot fi evaluaţi anumiţi indivizi din grup în faza de preintervenţie şi alţi indivizi în partea de postintervenţie (sau o combinaţie

între cele două: unii indivizi testaţi repetat, alţii testaţi o singură dată sau doar în anumite secvenţe). Acest lucru este posibil doar cu condiţia ca, în fiecare etapă de evaluare, selecţia indivizilor supuşi testării din cadrul unui cluster să fie una aleatorie.

La aceste avantaje aş adăuga şi posibilitatea mai ridicată de a utiliza experimente în mediu natural (ecologic) în cazul alocării randomizate la nivel de cluster (în grupurile lor naturale: echipe, clase etc.), decât dacă alocarea randomizată s-ar realiza la nivel individual, cu consecinţe benefice asupra validităţii externe a studiului.

Dincolo de aceste aspecte, randomizarea la nivel de cluster are însă şi o serie de dezavantaje comparativ cu cea realizată la nivel individual (Torgerson şi Torgerson, 2008), îndeosebi în ceea ce priveşte: (a) numărul de participanţi la studiu (pentru a asigura aceeaşi putere statistică, acestea au nevoie, de regulă, de un număr de participanţi la studiu mai mare cu un procent cuprins între 50% şi 200% decât în cazul randomizării la nivel individual, în funcţie de nivelul coeficientului de corelaţie intracluster – ICC); (b) apariţia aceloraşi probleme identificate în cazul utilizării metodei Zelen, ca urmare a alocării participanţilor în grupele experimentale înainte de a obţine consimţirea informată a participanţilor (deoarece adesea alocarea randomizată a clusterelor se face înaintea recrutării şi obţinerii consimţirii informate din partea participanţilor); (c) metodele statistice complexe de analiză a ipotezelor, astfel încât să ţină seama de organizarea multinivelară (ierarhică) a datelor; (d) posibilitatea redusă de mascare a intervenţiei primite, în general, de implementare a unor designuri de tip orb, precum şi un risc mai ridicat de distorsiune în evaluarea iniţială în privinţa aprecierii gradului de îndeplinire a condiţiilor de eligibilitate a participanţilor pentru a participa la studiu (ca urmare a cunoaşterii grupului din care aceştia vor face parte înainte de evaluarea

pretest); (e) restricţii de ordin logistic; clusterele alocate randomizat ar trebui să fie minimum patru pentru fiecare grup experimental (Murray, 1998) sau chiar şapte (Donner şi Klar, 2000), pentru a asigura prin anumite tehnici de alocare randomizată precum minimizarea, şansa distribuirii echilibrate la nivel de preintervenţie în cele două grupe experimentale. Această ultimă condiţie face ca experimentul cu alocare randomizată la nivel de cluster să fie mai puţin utilizat în realitate. De exemplu, dacă clusterul face referire la o clasă din şcoală, iar studiul îşi propune să compare eficienţa metodei A de predare comparativ cu metoda B, ar trebui să fie incluse în studiu minimum opt clase – câte patru clase pentru fiecare metodă, conform lui Murray (1998), respectiv minimum 14 clase – câte şapte pentru fiecare metodă, conform lui Donner şi Klar (2000).

În ciuda acestor limite, dacă este implementat corect, studiul randomizat controlat (la nivel de cluster) oferă o calitate şi o încredere superioară în rezultatele obţinute, faţă de oricare procedeu de tip cvasiexperimental. De aceea, recomandarea noastră ar fi ca, ori de câte ori nu este posibil să se apeleze la un studiu experimental cu alocarea randomizată individuală a participanţilor, studiul experimental cu alocare randomizată la nivel de cluster să constituie prima variantă alternativă de luat în calcul, pentru a asigura un nivel ridicat de validitate internă şi externă pentru concluziile obţinute în urma studiului.

4.2.3. Identificarea mecanismelor schimbării

Din punct de vedere statistic, modalitatea tipică prin care se pot identifica mecanisme ale schimbării responsabile pentru justificarea unei relaţii de tip cauză-efect pare a fi redusă la analiza de mediere. Prin intermediul

acestei analize se răspunde la un set de întrebări precum: (a) „intervenţia conduce la efectul dorit?” (relaţie existentă între VI şi VD); (b) „intervenţia conduce la modificări în nivelul variabilei mediatoare?” (relaţia existentă între VI şi VM); (c) „există o asociere statistică între VM şi VD?”; (d) „relaţia dintre VI şi VD este redusă sau devine absentă dacă controlăm statistic impactul VM asupra VD?” (Baron şi Kenny, 1986)[14].

Din punct de vedere metodologic însă, demonstrarea relaţiei statistice de mediere este necesară, dar nu şi suficientă, deoarece lipseşte raportarea temporală, aceea de a arăta că modificările la nivel de VD nu precedă, ci succedă modificărilor observate la nivel de VM, ca urmare a implementării intervenţiei.

Tabelul 4.8. *Tipul de design şi impactul acestuia asupra identificării mecanismelor schimbării*

	Mecanismul schimbării Evaluarea mediatorilor			Relaţia cauză-efect Evaluarea VD		
Tip de design	Pre	Interim	Post	Pre	Interim	Post
A. Măsurători concurente (standard)	X	O	X	X	O	X
B. Măsurători intermediare unice	X	X	X	X	X	X
C. Măsurători intermediare multiple	X X, X, X… X			X X, X, X… X		

Legendă: X – evaluare; O – absenţa evaluării

Kazdin (2006) oferă o discuţie elocventă cu privire la rolul designului unui studiu din perspectiva numărului de evaluări şi a aspectelor evaluate (atât VD, cât şi VM), pe care o vom simplifica (vezi tabelul 4.8) la cazul a trei designuri specifice întâlnite frecvent în studii randomizate.

În designul A, potenţialele variabile mediatoare sunt măsurate la începutul şi la finalul intervenţiei, împreună cu variabila dependentă de

interes. Nu există nici măsurători ale VD şi nici ale VM pe parcursul derulării intervenţiei. Un astfel de design, deşi ar permite statistic efectuarea unei analize de mediere, nu reuşeşte să diferenţieze temporal evoluţia datelor, adică nu permite să arate dacă modificările în VM precedă temporal modificările observate la nivel de VD, dacă situaţia este inversă sau dacă atât VM, cât şi VD sunt dependente de o altă variabilă (mecanism al schimbării), care nu a fost măsurată (ambele fiind efecte ale unei variabile străine terţe).

Designul B, spre deosebire de cazul A, include măsurarea VD şi a VM supuse investigării la un anumit moment pe parcursul implementării intervenţiei (de exemplu, după ce au fost parcurse jumătate din numărul de şedinţe prestabilite ale intervenţiei). Un asemenea design permite investigarea relaţiei dintre VD şi VM din perspectivă temporală (de exemplu, mai întâi se constată modificări ale nivelului VM, iar apoi se constată modificări la nivelul VD). În mod similar, se poate vedea în ce măsură magnitudinea modificărilor la nivelul VM (între preintervenţie şi evaluarea intermediară) prezice magnitudinea modificărilor la nivelul VD (între evaluarea intermediară şi evaluarea postintervenţie). Dezavantajul acestui design este determinat de nivelul ridicat al dependenţei (validitate a concluziilor) de momentul ales pentru evaluarea intermediară. Este posibil ca, la majoritatea participanţilor, modificările la nivel de VM să apară abia după parcurgerea a 80% din numărul de şedinţe prestabilite, fapt care va conduce la nesesizarea impactului VM asupra VD dacă momentul de evaluare al eficienţei intervenţiei a fost stabilit după ce au fost parcurse doar 50% dintre şedinţe. De aceea, varianta ideală din punct de vedere metodologic este aceea de a include mai multe puncte intermediare de evaluare (designul C).

Din punct de vedere statistic (de exemplu, dacă se apelează la ecuaţii structurale ca procedură de analiză statistică [*latent growth modeling*]) ar fi nevoie de minimum trei evaluări consecutive, atât ale evoluţiei VD, cât şi ale VM studiate, ceea ce ar însemna o singură evaluare intermediară pe lângă cele efectuate la nivel de pre-, respectiv postintervenţie. Pentru a surprinde relaţii complexe (inclusiv de tip neliniar) ale evoluţiei efectului ar fi preferabil să existe cât mai multe evaluări intermediare (de pildă cel puţin trei), fiind puţin fezabil ca evaluările să aibă loc după fiecare şedinţă de intervenţie (Singer şi Willett, 2003; Kazdin, 2006).

În sfârşit, Kazdin (2006) oferă o serie de recomandări de urmat de către cei interesaţi să identifice mecanisme ale schimbării responsabile pentru observarea unei relaţii de tip cauză-efect. Din această perspectivă, procesul identificării mecanismelor schimbării este unul multifazic. Într-o primă etapă, se recomandă măsurarea concomitentă a mai multor variabile cu potenţial mediator pentru a extrage într-o manieră competitivă acel mecanism care reuşeşte să explice cel mai mult din varianţa relaţiei dintre VI şi VD în cadrul unui design de tip C, cu măsurători multiple ale VM şi VD pe parcursul implementării intervenţiei. Mecanismele care trec de acest filtru (a. mediind statistic relaţia dintre VI şi VD mai bine decât o fac alte VM concurente; b. într-un context în care modificările la nivel de VM le precedă pe cele la nivel de VD), pot fi supuse ulterior testului direct. Acesta constă în manipularea experimentală a VI în direcţia varierii gradientul VM oferit. Aceste VM selectate trebuie să fie relevante teoretic şi sprijinite de dovezi empirice anterioare (vezi tabelul 4.9).

Tabelul 4.9. *Criterii utilizate în identificarea mecanismelor schimbării (Kazdin, 2007)*

Criteriu	**Explicitarea criteriului**
A. *Mediere*	Variabila considerată drept mecanism al schimbării ar trebui să fie identificată ca o variabilă mediatoare între variabila independentă şi cea dependentă.

B. *Specificitate*	Se distinge clar efectul specific al variabilei considerate drept mecanism al schimbării faţă de efectul altor variabile concurente (explicaţii alternative) sau al unor variabile confundate.
C. *Temporalitate*	Schimbările observate la nivelul variabilei mediatoare precedă temporal pe cele observate la nivelul variabilei dependente.
D. *Efect direct*	Manipularea experimentală a variabilei identificate ca mecanism al schimbării se soldează cu efecte în direcţia aşteptată la nivelul variabilei dependente.
E. *Consistenţă*	Dovezi convergente în sprijinul variabilei considerate drept mecanism al schimbării provin din studii variate în ceea ce priveşte contextul sau strategia de cercetare.
F. *Plauzibilitate*	Mecanismul schimbării selectat (identificat) este unul plauzibil, în acord cu cadrul conceptual (teoria) existent în domeniu.

Dintre criteriile enunţate mai sus, primele patru (A-D) constituie criterii minime necesare pentru a considera o anumită variabilă ca fiind un mecanism al schimbării.

4.3. Cvasiexperimentul – o strategie alternativă onorabilă

În realitate există foarte multe situaţii în care condiţiile restrictive necesare derulării unui experiment nu sunt îndeplinite. Cel mai adesea, piedica principală întâlnită în derularea unui experiment constă în imposibilitatea (caracterul nefezabil al) randomizării participanţilor din studiu în grupul experimental, respectiv în cel de control, indiferent de tipul de randomizare implicat. În spatele acestei dificultăţi de a randomiza participanţii în condiţiile experimentale se pot regăsi motive de ordin etic (de exemplu, nu ar fi deontologic să împărţim randomizat participanţii dintr-un studiu care ar viza efectele consumului de marijuana asupra stării

de sănătate, în grupe care consumă droguri uzual, care consumă acest drog ocazional, respectiv în grupa celor care nu consumă droguri). O altă categorie de motive care impietează asupra derulării unui experiment se referă la aspecte de ordin practic. De exemplu, pentru a compara eficienţa a două metode de predare este extrem de dificil, dacă nu imposibil, să se formeze clase cu o compoziţie aleatorie de elevi. La fel de dificil este să apelăm la zece clase paralele, pe care să le supunem unei randomizări la nivel de cluster. Însă este mult mai uşor să se compare eficienţa celor două metode prin implementarea cercetării în două clase paralele: a VII-a A – prima metodă, respectiv a VII-a B pentru cea de a doua metodă.

În contextul în care manipularea VI de către experimentator este posibilă, dar fără a exercita un control strict al variabilelor străine prin randomizarea participanţilor în situaţiile experimentale, vorbim despre un studiu cvasiexperimental. Acesta împărtăşeşte cu studiile experimentale clasice ideea controlării unor variabile de interes (variabila independentă) prin manipularea acesteia, însă diferă prin absenţa unui control adecvat al unor variabile străine, pe care randomizarea participanţilor în grupele experimentale le-ar fi rezolvat[15]. Mai mult, unele forme incipiente de studii cvasiexperimentale nici măcar nu apelează la ideea unui grup de control, ceea ce complică şi mai mult interpretarea rezultatelor prin riscurile existente la nivelul validităţii interne a concluziilor studiului.

O altă strategie de cercetare similară cu aceea a derulării unui studiu cvasiexperimental este aceea a derulării unui experiment natural. Deşi titulatura acestui tip de studii este una înşelătoare, merită remarcat că experimentele naturale sunt mult mai apropiate de ideea unui cvasiexperiment decât de aceea a unui experiment propriu-zis. În experimentele naturale nu există randomizarea participanţilor la studiu în grupe de comparare, iar manipularea variabilei independente, deşi există,

este declanşată natural, fără intervenţia experimentatorului. Un exemplu de experiment natural ar putea fi studierea impactului pe care trecerea în şomaj îl are asupra anxietăţii indivizilor. Pentru a studia această temă este posibil să se măsoare anxietatea indivizilor din cadrul unei companii după disponibilizarea a 50% dintre angajaţi, toţi având acelaşi profil ocupaţional – operatori de bandă. În acest caz există două grupe (angajaţi şi disponibilizaţi) constituite prin intervenţia altor forţe decât decizia experimentatorului, iar nivelul de anxietate este evaluat pentru a urmări dacă disponibilizarea a condus la observarea unui nivel de anxietate mai ridicat la angajaţii disponibilizaţi decât la cei păstraţi în companie.

Problema comună a tuturor designurilor cvasiexperimentale comparativ cu a celor experimentale este aceea a riscului ridicat de apariţie a distorsiunii de selecţie (a neechivalenţei grupurilor), cu consecinţe negative asupra validităţii interne a rezultatelor. De pildă, este posibil ca în grupul celor 50% de angajaţi disponibilizaţi să existe din start un număr mai mare de persoane cu un nivel de anxietate mai ridicat ca urmare a rezultatelor mai puţin bune înregistrate la ultima evaluare a performanţei în muncă, iar aceştia să fie în principal vizaţi pentru disponibilizare. Astfel, diferenţele sesizate pot să nu fie generate de acţiunea de disponibilizare, ci să fie determinate de diferenţele existente între cele două grupe la nivelul anxietăţii încă înainte de disponibilizare. În mod similar, compararea eficienţei a două metode de predare prin aplicarea metodei A de predare celor din clasa a IX-a A, respectiv a metodei B de predare celor din clasa a IX-a B, poate fi distorsionată dacă există din start diferenţe calitative între cele două clase (de exemplu, cei din clasa a IX-a A obţin performanţe mai bune nu pentru că metoda A de predare ar fi mai eficientă decât B, ci pentru că în clasa respectivă se găsesc mai mulţi elevi silitori decât în clasa paralelă a IX-a B). De aceea,

o măsură întâlnită frecvent în cazul studiilor cvasiexperimentale este aceea de a realiza evaluări atât înainte de implementarea intervenţiei (pretest), cât şi la finalul intervenţiei (posttest).

Dacă grupele nu diferă la nivel de pretest în ceea ce priveşte nivelul variabilei dependente măsurate, există premisele considerării celor două grupe (experimentală şi de control) ca fiind echivalente. Trebuie însă precizat că absenţa diferenţelor semnificative la nivel de VD cu ocazia evaluării preintervenţiei este o condiţie necesară, dar nu şi suficientă pentru a elimina riscul unei distorsiuni în selecţia participanţilor. Pentru a explicita această afirmaţie, voi apela la un ipotetic studiu cvasiexperimental privind efectul muzicii clasice asupra reducerii distresului perceput. Cercetătorul a dorit să vadă dacă cei care ascultă zilnic, timp de 10 zile, un calup de 30 de minute de muzică clasică atent selectată vor avea la final un nivel mai redus al stresului în comparaţie cu cei care urmează, în acelaşi interval de timp, şedinţe de lectură (beletristică). În studiul respectiv, cercetătorul nu apelează însă la randomizarea participanţilor în cele două grupe. Acestea sunt constituite după preferinţele participanţilor, în sensul că mai întâi sunt identificaţi 30 de indivizi care doresc să participe la şedinţele de meloterapie (fiind melomani), iar restul de 30 de indivizi rămaşi sunt distribuiţi automat în cealaltă grupă, de biblioterapie, considerată a fi grupa de control. Deşi există o evaluare înainte de începerea intervenţiei care nu identifică diferenţe semnificative statistic la nivelul stresului perceput, dar identifică asemenea diferenţe la nivel de posttest, nu este sigur în ce măsură diferenţele la nivelul stresului dintre cele două grupe sunt generate de eficienţa sporită a muzicii clasice sau de diferenţele în motivaţie existente între cele două grupe (de exemplu, indivizii din grupa celor care ascultă muzică clasică sunt acolo pentru că şi-au dorit [s-au autoselectat], în timp

ce indivizii distribuiţi în grupa de biblioterapie sunt acolo pentru că doar în grupa respectivă erau locuri disponibile).

Prin urmare, deşi existenţa unei evaluări la nivel de pretest este necesară pentru că poate identifica cazurile clare de grupe neechivalente ce pot afecta concluziile unui studiu, absenţa unei diferenţe semnificative la nivel de pretest privind VD nu reprezintă şi garanţia că cele două grupe sunt echivalente. Acest lucru se întâmplă pentru faptul că, deşi nivelul VD este unul echivalent la nivel de pretest, pot exista diferenţe între grupe cu privire la alte variabile importante (variabile confundate) ce pot influenţa evoluţia ulterioară a modificărilor la nivelul VD. De exemplu, grupa experimentală poate diferi de grupa de control la nivel motivaţional (nivelul de implicare şi al aşteptărilor, în special dacă participanţii sunt lăsaţi să-şi aleagă ei înşişi condiţia din care să facă parte) sau pot diferi cu privire la anumiţi factori de risc în perpetuarea unei stări de fapt nedorite. Pentru acest ultim caz, un exemplu ar fi situaţia a două departamente dintr-o organizaţie incluse într-un studiu care vizează impactul coaching-ului asupra satisfacţiei în muncă. Deşi acestea nu diferă la nivel de preintervenţie în privinţa nivelului satisfacţiei în muncă (VD), pot exista diferenţe între departamente cu privire la alţi factori cu impact asupra evoluţiei satisfacţiei în muncă, cum ar fi stilul de leadership exercitat de şeful de departament, autoritar într-un caz, democratic în celălalt.

Distorsiunea de selecţie a participanţilor nu este singurul risc întâlnit în cazul studiilor cvasiexperimentale. În funcţie de designul de cercetare ales, pot fi întâlnite şi alte riscuri la adresa validităţii interne. Shadish, Cook şi Campbell (2002) sintetizează principalele probleme inerente diverselor tipuri de studii cvasiexperimentale. O parte dintre aceste tipuri de design, considerate a fi mai frecvent întâlnite sau mai importante din punct de vedere a impactului lor asupra validităţii interne a concluziilor

unui studiu sunt sintetizate în tabelul de mai jos, fiind însoţite fiecare de un exemplu ipotetic. Precizăm însă că o detaliere a tuturor riscurilor majore privitoare la validitatea internă a unui studiu este prezentată în capitolul 3.

Tabelul 4.10. *Tipuri populare de design cvasiexperimental şi probleme asociate acestora*

Denumire, descriere şi exemplificare	**Avantaje şi/sau riscuri *majore specifice* designului ales**
Design pretest posttest fără grup de control **O1 X O2**[16] Un studiu care testează eficienţa unui program educaţional de combatere a fumatului la adolescenţi. 50 de adolescenţi sunt testaţi cu privire la nivelul lor de cunoştinţe cu privire la riscurile determinate de consumul de tutun, înainte şi după finalizarea şedinţelor de instruire. Nu există un grup de control.	***Avantaje*** Nu există (design nerecomandat din cauza multiplelor riscuri existente, dar prezentat datorită uşurinţei de implementare – de pildă orice program de intervenţie poate fi evaluat sumar prin evoluţia rezultatelor înainte şi după implementarea sa). ***Riscuri*** *Istoricul*. În absenţa unui grup de control nu este clar dacă efectul pozitiv se datorează programului respectiv sau altor aspecte neprevăzute, cum ar fi o lecţie de dirigenţie despre riscurile consumului de tutun. *Maturarea*. Chiar în condiţiile în care nu apar evenimente neprevăzute care să afecteze major efectul programului, este posibil ca, odată cu trecerea timpului şi expunerea repetată la mesaje din mass-media, nivelul de informaţii să fie mai bun la posttest decât la pretest. *Retragerea participanţilor*. Este posibil că la nivel de posttest să participe doar 25 (50%) dintre subiecţii iniţiali din studiu, restul nefiind interesaţi să finalizeze programul. În acest context ar exista o distorsiune spre pozitiv a rezultatelor, deoarece numai subiecţii motivaţi sunt luaţi în analiză (care deţin evaluări pre- şi posttest). *Testarea.* Este posibil ca nivelul VD să fie afectat ca urmare a testării repetate (de exemplu, performanţa participanţilor la testul de cunoştinţe despre fumat să fie influenţată de faptul că participanţii îşi aduc aminte răspunsurile date la testul grilă la nivel de pretest sau se

	interesează de unele răspunsuri la întrebările din pretest care le-au atras atenţia).
Design cu grup de control fără pretest (cu grupe neechivalente) **X O1** **O1** Acelaşi program educaţional de combatere a fumatului este testat apelând la două grupe (elevii din clasa a IX-a A beneficiază de program, în timp ce elevii din clasa a IX-a B nu beneficiază de program, fiind grupul de control). Evaluarea cunoştinţelor despre fumat se face o singură dată, la finalul intervenţiei.	***Avantaje*** Prin introducerea unui grup de control se poate răspunde la întrebarea: „ce s-ar fi întâmplat dacă intervenţia nu ar fi fost implementată?” (cât de bune sunt cunoştinţele adolescenţilor despre efectele fumatului dacă nu ar fi beneficiat de programul educaţional respectiv?). De asemenea, se elimină riscul legat de testarea repetată (familiarizarea cu sarcina). ***Riscuri*** *Distorsiunea de selecţie.* În absenţa randomizării şi în absenţa verificării la nivel de pretest a echivalenţei celor două grupe din perspectiva cunoştinţelor despre fumat, există riscul ca eventualele diferenţe dintre grupuri sesizate la nivel de posttest să fi existat încă de la început (de exemplu, clasa a IX-a A să aibă mai mulţi elevi buni la învăţătură decât cei din a IX-a B). O soluţie pentru a reduce riscul distorsiunii de selecţie constă în *potrivirea* participanţilor (*matching*), care în cazul respectiv ar însemna alegerea a două clase de elevi similare din perspectiva nivelului lor academic şi a profilului lor sociodemografic). Strategia de potrivire poate fi aplicată atât la nivel global (clasa ca unitate de măsură), cât şi la nivel individual (fiecărui individ inclus i se caută o pereche pentru a fi inclus în celălalt grup, cel de control). *Contaminarea.* E posibil ca în liceul respectiv elevii din cele două clase să se întâlnească frecvent cu diferite ocazii sociale, situaţii în care împărtăşesc între ei informaţii, inclusiv cele provenite din programul educaţional. Astfel, şi cei care nu au formal acces la informaţiile prezentate în program sunt contaminaţi de intervenţie, nerespectând cerinţele unui grup de control. O altă formă de contaminare posibilă este demoralizarea, respectiv ambiţionarea participanţilor din grupul de control, dacă aceştia cunosc faptul că nu sunt beneficiarii unei intervenţii de interes pentru ei.
Design cu grup de control ales pe	***Avantaje***

baza similarităţii (*matching*) **X O1** **O1** Deşi este tot un design cu grup de control fără pretest, în acest caz în alcătuirea grupelor se iau măsuri active ca cele două să tindă a fi grupuri echivalente. Un exemplu legat de programul educaţional de	Prin măsurile active luate în constituirea celor două grupe comparate se tinde spre a se obţine echivalenţa grupelor, astfel încât eventualele diferenţe dintre grupe să fie puse pe seama elementului diferit (intervenţia de care beneficiază o singură grupă). Pentru a obţine cele mai bune rezultate în aplicarea strategiei de potrivire se recomandă selectarea exclusivă a unor variabile cu nivel crescut de stabilitate cum ar fi aspectele demografice (vârstă, sex, nivel de studii) sau utilizarea unui scor compozit (*propensity score*) a cărui metodologie de calcul este disponibilă gratuit într-un format prietenos în MCHP (2012) şi explicată în secţiunea 5.5. Pentru o funcţionare optimă a acestei strategii de potrivire ar fi preferabil să se utilizeze loturi cu un număr mai mare de participanţi.
combatere a fumatului ar fi acela în care în compoziţia celor două grupe (de control, respectiv experimental) să intre adolescenţi similari (aceeaşi vârstă şi distribuţie de gen, aceeaşi expunere la fumat, proveniţi din şcoli similare ca performanţă academică şi profil sociocultural etc.), astfel încât singura diferenţă relevantă între cele două grupe, pentru a aprecia nivelul cunoştinţelor lor despre fumat, să fie aceea că unii beneficiază de programul educaţional, iar ceilalţi nu.	***Riscuri*** *Distorsiunea de selecţie.* Din păcate, acest demers are aceeaşi limită specifică logicii inductive. Astfel este posibil să existe şi alte diferenţe relevante între grupele comparate decât cele luate în calcul. Neincluderea lor printre criteriile de eligibilitate utilizate în selectarea participanţilor la studiu şi, mai ales, în constituirea celor două grupe, conduce la un risc crescut de absenţă a echivalenţei grupurilor şi, prin urmare, la un risc de a confunda diferenţele observate la nivel de posttest ca fiind datorate intervenţiei, în condiţiile în care ele sunt cauzate de distorsiunea de selecţie.
Design cu grup de control, cu pretest şi posttest **O1 X O2** **O1 O2** Programul educaţional de combatere a fumatului este evaluat prin intermediul a două grupe (una experimentală care urmează programul, respectiv una de control care nu urmează programul).	***Avantaje*** Prin introducerea unei evaluări la nivel de pretest se poate estima echivalenţa grupurilor (dacă cele două grupe nu diferă înainte de introducerea intervenţiei) sau se pot controla statistic eventualele diferenţe observate (dacă cele două grupe diferă la nivel de VD încă dinainte de implementarea intervenţiei), astfel încât să se reducă posibilitatea ca eventualele diferenţe sesizate la momentul postintervenţiei să fie o urmare a diferenţelor existente încă dinainte de activarea (manipularea) VI. ***Riscuri***

Evaluarea are loc atât înainte de derularea intervenţiei, cât şi după finalizarea acesteia.	*Contaminarea* (vezi explicaţia oferită anterior) *Interacţiunea dintre selecţie şi maturare*. Chiar dacă cele două grupe nu diferă la nivel de VD cu ocazia măsurătorii de tip pretest, acest lucru nu este o garanţie pentru absenţa unei diferenţe calitative între cele două grupe ca urmare a unei distorsiuni de selecţie. Aceasta se poate manifesta în timp prin ritmuri diferite de evoluţie, care nu se datorează intervenţiei în sine, ci unor caracteristici intrinseci ale grupurilor care îi fac să evolueze diferit. De exemplu, e posibil ca grupul experimental să fie constituit din adolescenţi cu un nivel intelectual mai ridicat. În asemenea cazuri, ritmul de îmbunătăţire a cunoştinţelor despre fumat poate fi mai mare în grupul experimental decât în cel de control, nu datorită intervenţiei, ci ca urmare a interacţiunii dintre selecţie şi maturare (de exemplu, copiii cu performanţe şcolare mai bune acumulează mai uşor cunoştinţe, inclusiv despre fumat).
Design cu grup de control, cu dublu pretest şi un posttest **O1 O2 X O3** **O1 O2 O3** Programul educaţional descris anterior este ajustat în sensul că nu există doar o singură evaluare la nivel de pretest, ci două asemenea evaluări, una cu trei luni înainte de demararea intervenţiei, iar cea de-a doua cu o săptămână înaintea începerii intervenţiei. Sunt evaluaţi atât cei din grupul experimental, cât şi cei din grupul de control.	***Avantaje*** Spre deosebire de designul anterior, prin compararea evoluţiei de la O1 la O2 cu cea observată în grupul experimental de la O2 la O3, se poate observa mai bine dacă eventualele diferenţe observate sunt cauzate de intervenţia aplicată sau apar natural din cauze menţionate precum istoricul sau maturarea. Astfel, dacă asumăm că efectele maturării pot avea acelaşi ritm între O1 şi O2, ca între O2 şi O3, intensificarea ritmului de creştere observat între O2 şi O3 se datorează mult mai probabil efectelor intervenţiei. În ciuda acestor avantaje, designul este dificil de implementat, deoarece participanţii la studiu ar trebui să aştepte destul de mult până să primească intervenţia necesară (distanţa temporală dintre O1 şi O2 fiind egală cu cea dintre O2 şi O3). ***Riscuri*** *Contaminarea* (vezi explicaţia de mai sus) *Testarea* (vezi explicaţia de mai sus)
Design cu grup de control şi inversare de rol **O1 X O2 O3**	***Avantaje*** Prin introducerea inversării de rol între grupele incluse în studiu se obţine un design puternic din perspectiva

O1 O2 X O3 Cele două clase participante sunt evaluate iniţial cu privire la nivelul lor de cunoştinţe despre riscurile fumatului. Programul este implementat mai întâi în clasa a IX-a A, în timp ce a IX-a B serveşte drept grup de control. La finalul intervenţiei cele două clase sunt evaluate din nou. Apoi intervenţia este aplicată celor din clasa a IX-a B, în timp ce clasa a IX-a A preia rolul de grup de control. Ultima evaluare pentru ambele clase are loc după finalizarea intervenţiei la clasa a IX-a B.	validităţii interne a studiului. Totuşi, designul nu constituie un demers de replicare a efectului intervenţiei deoarece grupul experimental iniţial nu este un grup clasic de control dat fiind că efectele determinate de intervenţia întreruptă sunt diferite deseori de cele observate la un grup de control clasic (fără intervenţie). ***Riscuri*** *Contaminarea* (îndeosebi *compensarea*). Ca urmare a întreruperii intervenţiei în grupul experimental între O2 şi O3 este posibil ca participanţii din grupa respectivă să încerce să compenseze întreruperea intervenţiei sau să devină demotivaţi de această întrerupere, comportându-se într-un mod diferit faţă de un grup de control clasic.
Design cu serii întrerupte, grup de control şi inversare de rol **O1 O2 X O3 O4 O5 O6 O7** **O1 O2 O3 O4 O5 X O6 O7** Caz ipotetic: poliţia din Serbia a introdus în 2007 măsura arestării conducătorilor auto care sunt surprinşi conducând autoturismul cu peste 30 de km/h peste limita legală. Aceeaşi măsură a fost implementată în Bulgaria în 2010. Date oficiale despre numărul de accidente soldate cu victime sunt colectate anual din cele două ţări din perioada 2004-2013. Se observă o scădere abruptă a numărului de victime din accidente de circulaţie după introducerea acestei măsuri.	***Avantaje*** Prin complexitatea observaţiilor colectate, acest tip de design este unul dintre cele mai robuste forme de design cvasiexperimental. Este o strategie potrivită de a analiza efectele la nivel sistemic produse de anumite decizii guvernamentale. Într-o variantă simplificată, cel de-al doilea grup poate servi doar drept grup de control (fără a implica vreo intervenţie). ***Riscuri*** *Testarea.* Ca urmare a numărului mare de măsurători repetate, acest design este mai puţin pretabil în cazul în care VD este măsurată prin autoraportări, fiind mai potrivit pentru a fi utilizat în situaţii cu VD măsurate obiectiv. *Efectele întârziate*. Unele modificări nu produc efecte aşteptate imediat, ci efecte întârziate, observabile într-o etapă ulterioară din şirul de observaţii (fiind, astfel, mai greu de intuit legătura cauzală potenţială). *Distorsiunea de selecţie*. Este puţin probabil să se atingă dezideratul echivalenţei între grupurile comparate, de aceea este posibil ca efectele introduse de intervenţie să nu se replice din cauza unor caracteristici specifice grupului (efectul introducerii tomberoanelor ecologice

	este dependent de mentalitatea colectivităţii în care este introdusă măsura). *Variabilele confundate*. Este greu de decelat cu claritate cauza modificărilor observate. În exemplul oferit este posibil ca în una dintre ţări modificarea să se fi suprapus cu înăsprirea amenzilor de circulaţie, iar în cazul celeilalte ţări, măsura să se fi suprapus cu o îmbunătăţire semnificativă a infrastructurii (500 de km de noi autostrăzi şi 300 de km de noi centuri ocolitoare). Ambele modificări au, la rândul lor, consecinţe benefice asupra scăderii numărului de accidente de circulaţie.
Designul cu linie întreruptă de regresie (*regression discontinuity design* – RDD) **O C > X O2** **O C ≤ O2** Acest tip de design se distinge de celelalte forme de studii cvasiexperimentale nu prin format (de exemplu, poate fi un design cu grup de control, cu evaluări pre- şi posttest), ci prin specificul împărţirii participanţilor în grupul experimental, respectiv în cel de control (în baza rezultatelor observate la pretest, mai exact a unui scor-limită [*cutoff score*]). Scorul de la pretest utilizat pentru divizare poate fi cel obţinut la VD, de la o altă variabilă, fie ea corelată sau nu cu VD sau poate fi chiar un scor compozit prin reunirea mai multor variabile. Un exemplu de acest tip de design este oferit în caseta 4.6.	***Avantaje*** În ciuda lipsei de inspiraţie în denumirea acestui design, acesta este considerat a fi cel mai puternic tip de design cvasiexperimental, fiind considerat de mulţi (Shadish *et al.*, 2002) la fel de eficient ca şi studiile experimentale în identificarea unei relaţii de tip cauză-efect, deşi nu apelează la randomizare ca modalitate de control a diferitelor variabile confundate. Este recomandat de specialişti (Torgerson şi Torgerson, 2008) spre a fi utilizat ori de câte ori nu este etic posibil să se apeleze la randomizare (studii experimentale). ***Riscuri*** *Puterea statistică scăzută*. Pentru a ajunge la aceeaşi putere statistică pe care o posedă un studiu de tip experimental, este nevoie de colectarea datelor de la un număr dublu sau chiar triplu de participanţi pentru a atinge aceeaşi putere statistică. Numărul de participanţi necesari poate fi chiar de 3-4 ori mai mare decât într-un studiu experimental, pe măsură ce scorul limită ales este mai îndepărtat faţă de medie/mediana variabilei respective, pe măsură ce variabila de grupare corelează mai mult cu VD. *Artefacte ce induc erori suplimentare*. Cunoaşterea publică a scorului-limită utilizat în decizia de a include participantul în grupul experimental sau în cel de control poate spori riscul unor evaluări subiective (o indulgenţă sau exigenţă sporită în evaluarea nivelului variabilei utilizate pentru a grupa indivizii). Aceeaşi problemă

	există şi în cazul alegerii unor scoruri limită care conduc natural la consecinţe diferite (de exemplu, utilizarea notei 5 ca standard minim pentru a promova un examen este asociată adesea cu situaţia în care mai mulţi indivizi obţin o asemenea notă decât una mai mică, incluzând aici cazurile în care profesorii acordă „din milă” o asemenea notă, inclusiv unora care nu îndeplineau criteriile minime pentru promovare). De aceea, indicat ar fi ca scorul-limită ales să nu fie identic cu anumite criterii oficiale care să conducă la consecinţe diferite (recompense *vs* penalizări).

Caseta 4.6. *O ilustrare a unui design cu linie întreruptă de regresie*

O asociaţie caritabilă organizează anual meditaţii gratuite pentru elevii ce provin din familii defavorizate cu scopul de a-i ajuta să promoveze bacalaureatul. Statul român se gândeşte să subvenţioneze 25% din cheltuielile asociaţiei legate de acest proiect dacă se dovedeşte că programul este unul eficient (îmbunătăţeşte performanţa academică a elevilor). Asociaţia caritabilă este de acord cu această propunere, însă sugerează că eficienţa programului de meditaţii ar trebui demonstrată altfel decât printr-un studiu experimental, deoarece ar fi lipsit de etică ca, în baza randomizării, o parte dintre tinerii aflaţi în grupul său ţintă (tinerii proveniţi din familii defavorizate) să nu beneficieze de această pregătire suplimentară pentru simplul fapt că au fost distribuiţi aleatoriu în grupul de control. De aceea, ei îşi propun să demonstreze eficienţa programului prin intermediul unui design cu regresie întreruptă. Pentru aceasta, în locul procedurii de randomizare, ei decid să împartă elevii în cele două grupe (experimentală şi de control) în baza unui criteriu ce vizează severitatea dificultăţilor materiale cu care se confruntă familiile elevilor (în ideea că cei mai nevoiaşi să beneficieze obligatoriu de acest program). Astfel, toţi elevii al căror venit net per membru de familie se situează sub 100 de euro/lună au fost cuprinşi în grupul experimental, iar cei care aveau un venit mediu per membru de familie peste 100 de euro/lună, dar sub media naţională erau cuprinşi în grupul de control. Astfel, suma de 100 de euro venit lunar/membru familie a reprezentat scorul-limită în baza căruia au fost constituite cele două grupe – experimentală şi de control (venitul fiind variabila de grupare). Conform acestui algoritm de împărţire au fost distribuiţi 120 de elevi în grupul experimental, respectiv 160 de elevi în grupul de control. Cei din grupul experimental au urmat şedinţe săptămânale de meditaţie timp de trei luni, în timp ce participanţii din grupul de control nu au urmat şedinţe de meditaţie nici în acest cadru organizat, nici prin intermediul altor programe sau demersuri individuale, bazându-se în pregătirea lor doar pe informaţiile învăţate la şcoală (pentru că resursele lor materiale nu permiteau apelul la strategii alternative, iar

această stare de fapt a fost verificată ulterior în baza unui interviu). Variabila dependentă urmărită a constat în media generală obţinută la bacalaureat, fiind vorba de un punctaj cu valori între 1.00 şi 10.00[17].

Concluzii

În acest capitol au fost cuprinse informaţii introductive despre principalele modalităţi de studiu ştiinţific a unei relaţii de tip cauză-efect. Experimentul rămâne calea regală de studiu a aspectelor psihologice, datorită caracteristicilor sale care îi permit să controleze cât mai mulţi factori străini (variabile confundate) şi să evite distorsiuni ce pot afecta concluziile studiilor. Din acest punct de vedere, studiile experimentale ar trebui să constituie prima linie de atac în abordarea ştiinţifică a oricărei probleme psihologice care a depăşit faza studiilor efectuate în scop explorator sau descriptiv. Cea mai mare parte a capitolului este destinată prezentării diferitelor tipuri de experimente, cu accent pe un tip particular de design specific intervenţiilor, denumit generic studiu randomizat controlat. Cu toate că cel mai răspândit tip de randomizare are loc la nivel individual, o alternativă posibilă este randomizarea la nivel de cluster.

Dată fiind importanţa acestei categorii de studii, am prezentat pe parcursul capitolului o serie de recomandări practice legate de implementarea unui studiu randomizat controlat, care vizează: recrutarea participanţilor la studiu, caracteristicile formularului de consimţire, implementarea studiului, etapa de urmărire şi analiza datelor (atât pentru a surprinde o relaţie de tip cauză-efect, cât şi pentru a identifica potenţiale mecanisme psihologice responsabile pentru schimbarea produsă).

Atunci când randomizarea nu este posibilă nici la nivel individual, nici la nivel de cluster, fie din motive etice, fie din motive ce ţin de

fezabilitatea implementării, recomandăm utilizarea unei strategii alternative, aceea de a implementa studii de tip cvasiexperimental. Acestea împărtăşesc cu studiile experimentale ideea manipulării variabilei independente (intervenţiei) de către cercetător pentru a vedea efectele acestei acţiuni asupra variabilei dependente, însă prezintă mai multe riscuri cu privire la validitatea internă a concluziilor extrase din cauza absenţei randomizării participanţilor în grupele experimentale, măsură care ar fi asigurat (cel puţin la nivel teoretic) un control al variabilelor confundate. Din multitudinea de designuri cvasiexperimentale sunt prezentate câteva dintre cele mai populare. Între acestea se distinge designul cu linie de regresie întreruptă, considerat a fi cel mai puţin afectat de riscuri la nivelul validităţii interne a studiului dintre designurile de tip cvasiexperimental.

Lecturi suplimentare

I. Boutron, D. Moher, D.G. Altman, K. Schulz, P. Ravaud, for the CONSORT group (2008). Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*, 148, pp. 295-309.

Sistemul CONSORT este probabil cel mai avansat model de raportare a studiilor experimentale, fiind menţionaţi şi explicaţi paşii necesari de urmat pentru redactarea unor articole de calitate fundamentate pe studii randomizate. În articolul menţionat accentul este pus pe o anumită categorie de intervenţii (cele nefarmaceutice), aplicabile cazurilor specifice din psihologie sau psihoterapie.

Sandra Eldridge, Sally Kerry (2012). *A Practical Guide to Cluster Randomised Trials in Health Services Research*. Chichester, UK: Willey & Blackwell.

Cititorii interesaţi să aprofundeze/implementeze studii randomizate controlate la nivel de cluster vor găsi aici un ghid aplicativ relevant, cu toate că exemplele oferite provin exclusiv din sfera sănătăţii.

Alan E. Kazdin (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3, pp. 1-27.

Articolul discută exigenţele metodologice şi statistice pentru a identifica mecanisme ale schimbării (variabile mediatoare), responsabile pentru modificările observate la nivelul variabilei dependente în urma apelului la psihoterapie. Deşi exemplele se rezumă la sfera psihologiei clinice, logica prezentată poate fi extinsă către alte tipuri de intervenţie, fie ele din mediul organizaţional sau educaţional.

Sebastian Pintea (2010). The relevance of results in clinical research: statistical, practical, and clinical significance. *Journal of Cognitive and Behavioral Psychotherapies*, 10(1), pp. 101-114, disponibil şi online la http://www.scipio.ro/documents/13115/2644c839-844b-4d6f-aaec-dee2095671f5

Articolul detaliază, pentru cei interesaţi, indicatorii de calcul ai eficienţei unei intervenţii din perspectivă clinică (client recuperat, ameliorat etc.). Aceste inovaţii nu sunt echivalente conceptului standard de mărime a efectului, chiar dacă sunt dependente indirect de acesta, fiind utile îndeosebi din perspectivă aplicativă, în comunicarea cu beneficiarii serviciilor sau cu cei care decid finanţarea din bani publici a unor asemenea servicii.

William R. Shadish, Thomas D. Cook, Donald T. Campbell (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, USA: Houghton Mifflin Company.

Dacă ar fi să nominalizez o singură carte însoţitoare a capitolului de faţă, cu o serie de texte extensive despre problematica experimentelor şi cvasiexperimentelor, atunci aceasta ar fi nominalizarea făcută. Cartea prezintă în detaliu avantajele şi riscurile existente pentru diferite designuri experimentale şi cvasiexperimentale discutate în acest capitol.

William Trochim (1990). Regression-discontinuity design in health evaluation. În L. Sechrest, E. Perrin şi J. Bunker (eds.). *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*. U.S. Dept. of HHS, Agency for Health Care Policy and Research, Washington, D.C. Disponibil online, alături de alte articole pe aceeaşi temă, la: http://www.socialresearchmethods.net/research/rd.htm

Articolul respectiv oferă o introducere detaliată asupra caracteristicilor unui design cu linie de regresie întreruptă, inclusiv cu referire la paşii caracteristici de urmat în analiza statistică a datelor provenite dintr-un asemenea design.

Bibliografie

Barlow, D.H., Nock, M.K., Hersen, M. (2008). Single case experimental designs: Strategies for studying behavior change (ediţia a III-a). Boston: Allyn & Bacon.

Baron, R.M., Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of*

Personality and Social Psychology, 51, pp. 1173-1182.

Bjelakovic, G., Nikolova, D., Gluud, L.L. *et al.* (2007). Mortality in randomized trials of antioxidant supplements for primary and secondary prevention: systematic review and metaanalysis. *Journal of the American Medical Association*, 297, pp. 842-857.

Bradt, J., Dileo, C. (2009). Music for stress and anxiety reduction in coronary heart disease patients. *Cochrane Database of Systematic Reviews*, 15, CD006577.

Cordray, D.S. Pion, G.M. (2006). Treatment strength and integrity: Models and methods. În R.R. Bootzin P.E. McKnight (eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103-124). Washington, DC: American Psychological Association.

Donner, A. Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. Londra: Arnold.

Elmes, D.G., Kantowitz, B.H., Roediger III, H.L. (2012). *Research Methods in Psychology* (ediţia a IX-a). Wadsworth/Thomson Learning, Inc.

Gherman, G. (2008). *Impactul muzicii asupra anxietăţii post-operatorii la pacienţii cu intervenţii de stentare*. Lucrarea de licenţă la Universitatea de Vest din Timişoara. Manuscris nepublicat.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., Crombez, G. (2010). Evaluative conditioning in humans: A metaanalysis. *Psychological Bulletin*, 136, pp. 390-421.

Kazdin, A.E. (2006). Arbitrary metrics: Implications for identifying evidence based treatments. *American Psychologist*, 61(1), pp. 42-49.

Kazdin, A.E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3, pp. 1-27.

Kazdin, A.E. (2010). *Single-case Research Designs: Methods for Clinical and Applied Settings* (ediţia a II-a). New York: Oxford University Press.

Klassen, J.A., Liang, Y., Tjosvold, L., Klassen, R.P., Hartling, L. (2008). Music for pain and anxiety in children undergoing medical procedures: A systematic review of randomized controlled trials. *Ambulatory Pediatrics*, 8(2), pp. 117-128.

Lando, H.A. (1982). A factorial analysis of preparation, aversion, and maintenance in the elimination of smoking. *Addictive Behaviors*, 7, pp. 143-154.

Lilienfeld, S.O., Lynn, S.J., Ruscio, J., Beyerstein, B.L. (2010). *50 great myths of popular psychology*: *Shattering widespread misconceptions about human behavior.* Chichester, Anglia: Wiley-Blackwell.

McDonald, A.M., Knight, R.C., Campbell, M.K., Entwistle, V.A., Grant, A.M. *et al.* (2006). What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials*, 7, p. 9.

MCHP (Minitoba Centre for Health Policy) (2012). *Propensity score matching in observational studies.* Găsit la: http://umanitoba.ca/faculties/medicine/units/mchp/protocol/media/propensity_score_matching.pdf.

Moher D., Hopewell, S., Schulz, K.F., Montori, V., Gotzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., Altman, D.G. (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, 340, pp. 1-28.

Murray, D.M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Urbaniak, G.C., Plous, S. (2011). Research Randomizer (Version 3.0) [Computer software]. Găsit la: 24 August 2012, http://www.randomizer.org/

Pelham, B.W., Blanton, H. (2007). *Conducting research in psychology: Measuring the weight of smoke* (ediţia a III-a). Belmont, CA: Wadsworth/Thomson.

Pfister, M., Jones S.J., Cheek, D. (2011). Music Therapy for Preoperative Anxiety. Găsit la: http://www.crna.tcu.edu/docs/dnp-a%20capstone/Pfister%20M%20Use%20of%20Music%20to%20Minimize%20Preoperative%20Patient%20Anxiety%202011.pdf

Piaggio, G., Elbourne, D.R., Altman, D.G., Pocock, S.J., Evans, S.J. (2006). Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *Journal of the American Medical Association*, 295, pp. 1152-1160.

Pietschnig, J., Voracek, M., Formann, A.K. (2010). Mozart effect – Shmozart effect: A metaanalysis. *Intelligence*, 38, pp. 314-323.

Pintea S. (2010). The relevance of results in clinical research: statistical, practical and clinical significance. *Journal of Cognitive and Behavioral Psychotherapies*, 10(1), pp. 101-114.

Rauscher, F.H., Shaw, G.L., Ky, K.N. (1993). Music and spatial task performance. *Nature*, 365, p. 611.

Riegelman, R. (1979). Contributory cause: unnecessary and insufficient. *Postgraduate Medicine*, 66, pp. 177-179.

Sava, F.A., Rusu, S. (2011). Evaluarea personalităţii prin tehnica amorsării. În F.A. Sava (ed.), *Metode implicite de investigare a personalităţii* (pp 95-123). Iaşi: Editura Polirom.

Schulz, K.F., Altman, D.G., Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine*, 152(11), pp. 1-8.

Shadish, W.R., Cook, T.D., Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Singer, J.D., Willett, J.B. (2003). *Applied Longitudinal Data Analysis: Methods for Studying Change and Event Occurrence*. New York: Oxford University Press.

Torgerson, D. Torgerson, C. (2008). *Designing and running randomised trials in health, education and the social sciences*. Basingstoke: Palgrave Macmillan.

1. În limba engleză se regăseşte sub denumirea de *carry-over effect*. Acest tip de efect constituie un risc atunci când nu se urmăreşte explicit măsurarea acestui efect. Dacă se are în vedere surprinderea sau accentuarea acestui efect, precum în cazul tehnicilor de amorsare (vezi Sava şi

Rusu, 2011), atunci efectul nu constituie un risc, ci dimpotrivă constituie elementul activ căutat prin manipularea experimentală.

2. Studiile care vizează obiective cauzal-explicative sunt focalizate pe identificarea mecanismelor schimbării (de exemplu, a factorilor responsabili pentru eficienţa unui anumit tip de psihoterapie în diminuarea anxietăţii sociale). Aceste mecanisme joacă rolul unor variabile mediatoare între VI şi VD, contribuind la producerea unui lanţ cauzal (VI modifică mecanismul schimbării, iar acesta, la rândul lui, modifică nivelul VD).

3. În literatura de specialitate sunt mai multe riscuri privitoare la validitatea internă a acestui tip de design, o parte dintre acestea, precum efectul maturării, reactivitatea măsurării, istoricul, expectanţele experimentatorului (Shadish *et al.*, 2002) au fost prezentate în detaliu în capitolul anterior (în secţiunile dedicate validităţii interne şi celei de construct).

4. Unii cercetători definesc aceste tipuri de distorsiuni drept artefacte. În opinia noastră, ele reprezintă mai degrabă variabile confundate deoarece pot afecta validitatea internă a unui studiu, putând fi percepute ca nişte artefacte doar dacă consecinţele lor vizează afectarea altor tipuri de validitate (externă, de construct) şi dacă influenţa lor a fost una relativ constantă asupra tuturor celor investigaţi.

5. Shadish, Cook şi Campbell (2002) oferă sintaxa de lucru în SPSS, precum şi un macro de lucru în Excel pentru randomizare simplă. O altă resursă utilă ar fi accesarea de programe de randomizare disponibile online precum resursa pusă la dispoziţie de Urbaniak şi Plous (www.randomizer.org). Aceste resurse sunt utile pentru cazul în care alocarea randomizată se face simultan, însă pot fi adaptate şi pentru alocarea randomizată secvenţială.

6. De exemplu, în cazul psihoterapiei, participanţii într-un grup de control urmează acelaşi număr de şedinţe de psihoterapie, precum cei din grupul experimental. În aceste întâlniri ei discută într-o manieră nespecifică (ca într-o discuţie amicală), despre problema lor, potenţiale cauze ale acesteia, despre alte aspecte din viaţa lor şi despre posibile soluţii, însă fără a apela la strategiile specifice alocate grupului experimental.

7. Situaţiile prezentate în tabelul 4.3 prezintă doar cazurile simple, ale comparării a două grupe. În cazul studiilor experimentale bazate pe compararea a mai mult de două grupe, pot fi găsite variante combinate. De pildă, într-un studiu bazat pe trei grupe experimentale: (a) intervenţii cognitiv-comportamentale; (b) intervenţii familial-sistemice şi (c) intervenţie uzuală, vom putea urmări ca obiective atât să comparăm care dintre cele două forme de psihoterapie este mai eficientă, cât şi să vedem dacă efectul se datorează unor factori specifici acestor terapii, prin raportarea rezultatelor fiecărui tip de terapie la intervenţia uzuală (nespecifică). De asemenea, designurile pot fi complicate prin introducerea sau luarea în calcul a unor variabile moderatoare sau mediatoare.

8. Acest tip de studii este cunoscut şi sub denumirile de trialuri clinice, studii randomizate, studii randomizate clinice, denumiri care fac referire la sfera lor de aplicare cel mai frecvent întâlnită (medicină, farmacie, psihoterapie etc.; RCT – *randomized clinical trial*, *clinical trial*, *randomized trial*). Între RCT şi studii clinice controlate se face uneori o distincţie fină, în

sensul că ultimele nu presupun randomizarea participanţilor în condiţiile experimentale, fiind astfel incluse în zona studiilor cvasiexperimentale. Această nuanţă nu este însă utilizată în acest mod de către toţi autorii, unii dintre ei utilizând terminologia de studii clinice controlate pentru a se referi tot la un RCT.

9. Ar fi fost recomandat să se expliciteze succint ce resurse ar solicita includerea în grupul care beneficiază în plus de o intervenţie psihologică, cum ar fi numărul total de şedinţe sau frecvenţa acestora.

10. Termeni similari din literatura de specialitate care fac referire la livrarea acurată a intervenţiei sunt „integritatea” sau „fidelitatea intervenţiei” (*treatment integrity* sau *treatment fidelity*) (Cordray şi Pion, 2006).

11. O măsură întâlnită în alte studii experimentale decât cele focalizate pe intervenţii (de exemplu, în cazul manipulării instructajului oferit participanţilor sau a intensităţii stimulilor prezentaţi etc.) este de a testa existenţa efectului indus prin manipularea experimentală (*manipulation check*). Acest procedeu a fost descris în capitolul anterior, când ne-am referit la aspecte ale validităţii de construct legate de manipularea VI. Din perspectiva validităţii de construct, atât verificarea integrităţii intervenţiei, cât şi verificarea existenţei efectului indus prin manipulare, ne ajută să interpretăm corect rezultatele, din perspectiva implementării adecvate a VI.

12. Carroll *et al*. (1998 *apud* Cordray şi Pion, 2006) au evidenţiat că un procent cuprins între 28% şi 38% dintre participanţii incluşi în studii randomizate au apelat, timp de cel puţin o şedinţă, la alte servicii (intervenţii) decât cele care le-au fost alocate în studiu.

13. Logica unui experiment spune că, în urma alocării randomizate a indivizilor în condiţiile experimentale, există o probabilitate foarte mică ca grupele comparate să difere la momentul pretestului, astfel că diferenţele observate la momentul postintervenţiei sunt considerate ca fiind cauzate prin manipularea VI (de intervenţie). O excepţie de la această situaţie este aceea în care grupul experimental îşi modifica nivelul VD din pretest în posttest în direcţia dezirabilă, dar nesemnificativ statistic dacă comparăm magnitudinea schimbării VD din pretest în posttest, iar grupul de control îşi modifica VD din pretest în posttest în direcţia opusă, dar tot la un nivel nesemnificativ statistic. Este însă posibil ca analiza de la nivel de postintervenţie să indice diferenţe semnificative dintre grupuri. Dacă ne vom baza numai pe comparaţia intergrupală de la momentul postintervenţiei, vom afirma că intervenţia este eficientă, deşi, practic, modificarea VD din grupul experimental nu a fost una semnificativă statistic. Această problemă, rară în studiile experimentale, poate apărea frecvent în studiile cvasiexperimentale (în care nu este asigurată echivalenţa grupelor la nivel de VD în pretest), în cazul cărora analiza atât la nivel intergrupal, cât şi intragrupal, devine obligatorie. Mai mult, în studiile cvasiexperimentale se poate ajunge uneori la concluzia că o intervenţie este eficientă chiar dacă nu există diferenţe semnificative statistic la nivel de postintervenţie. De exemplu, într-un studiu în care se testează eficienţa unei noi metode de predare a statisticii la studenţi, se poate ajunge la concluzia că metoda este eficientă, chiar dacă ambele grupe înregistrează la posttest o medie a performanţelor în statistică egală cu 8,00, în condiţiile în care cei care au avut parte de

noua metodă de predare aveau la pretest media 6,00, în timp ce grupul de control, expus unei metode tradiţionale de predare, avea la pretest aceeaşi medie ca şi la posttest, adică 8,00.

14. Din perspectiva relaţiilor de mediere în explicitarea unor mecanisme cauzale se vorbeşte despre efecte directe (VM asupra VD) şi efecte indirecte (VI asupra VD). În terminologia unor relaţii de tip cauză-efect, putem vorbi despre cauze proximale (VM pentru VD), respectiv despre cauze distale (VI asupra VD), asumând că lanţul cauzal analizat are doar trei elemente (nu există un şir de variabile mediatoare între VI şi VD).

15. O alternativă interesantă pentru testarea unor relaţii cauzale atunci când studiile randomizate la nivel individual sau la nivel de cluster nu sunt fezabile din punct de vedere logistic este experimentul derulat pe un singur subiect (*single-case experiment*). În domeniul psihologiei clinice, efectuarea unei serii considerabile de experimente pe un singur subiect poate constitui o dovadă mai puternică din perspectiva validării ştiinţifice a intervenţiei respective decât cea obţinută dintr-un singur studiu de tip cvasiexperimental. Detalii suplimentare despre designul şi analiza experimentelor cu un singur subiect sunt oferite de Barlow, Nock şi Hersen (2008).

16. X – intervenţie; O1, O2 – observaţii (evaluarea efectelor)

17. În studiile experimentale şi cele cvasiexperimentale clasice suntem interesaţi să găsim diferenţe semnificative între mediile celor două grupe pentru a considera intervenţia ca fiind eficientă. În schimb, în designul liniei de regresie întrerupte se urmăreşte dacă linia de regresie dintre variabila cu rol de grupare şi VD are caracteristici diferite la nivel de intercept şi pantă de regresie (linie discontinuă în jurul valorii scorului-limită), inclusiv între indivizii similari – cei incluşi în grupul de control (cu un venit mediu de 101 euro/lună) sau alocaţi în grupul experimental (cu un venit mediu lunar foarte apropiat de 99 de euro lunar net/membru de familie).

Capitolul 5
Studiile non-experimentale

Una dintre afirmaţiile repetate frecvent în acest volum este aceea că ştiinţa avansează gradual. De exemplu, nu putem sări direct la a identifica relaţii cauzale între fenomene şi la a elucida mecanisme explicative pentru această stare de fapt, fără să fi găsit mai întâi că variabilele supuse analizei corelează între ele sau fără a explora iniţial fenomenul studiat. De asemenea, chiar dacă suntem interesaţi de găsirea unor relaţii de tip cauzal, uneori nu este posibilă derularea unor experimente sau cvasiexperimente din motive etice sau de fezabilitate. De exemplu, pentru a identifica impactul cauzal al statutului socioeconomic sau al genului persoanei asupra altor variabile de interes ar fi lipsit de etică sau imposibil să manipulăm variabilele amintite. De aceea, prima linie de abordare a oricărui fenomen psihic studiat este reprezentată, de obicei, de studiile de tip non-experimental. Acestea au drept caracteristică de bază faptul că măsurătorile au loc într-un context nemodificat (nemanipulat) de către cercetător, acest lucru fiind asociat frecvent cu ideea exercitării unui control redus asupra variabilelor cauzale de interes[1].

Spre deosebire de zona studiilor experimentale în cadrul căreia există un acord de clasificare a studiilor în două categorii majore (experimente şi cvasiexperimente), în cazul studiilor de tip non-experimental lucrurile nu stau la fel. De exemplu, se vorbeşte despre studii de caz, despre sondaje de opinie, despre studii corelaţionale, despre studii comparative *ex post facto*, despre studii longitudinale etc., situaţie în care sunt amestecate criterii diferite de definire. De altfel, Johnson (2001 *apud* Belli, 2009) concluziona după o trecere în revistă a nu mai puţin de 23 de cărţi de metodologia cercetării din

domeniul ştiinţelor sociale că nu există o poziţie unitară în modul de clasificare a diferitelor tipuri de studii non-experimentale, cele mai des menţionate fiind studiile corelaţionale, sondajele de opinie şi studiile comparative de tip *ex post facto*. Acelaşi autor (Johnson, 2001 *apud* Belli, 2009) încerca să clasifice studiile non-experimentale în funcţie de două criterii: cel al temporalităţii, respectiv cel al scopului pentru care era efectuată cercetarea.

După primul criteriu, s-ar putea vorbi despre studii retrospective, studii transversale (*cross-sectional*) şi despre studii prospective. În studiile retrospective, o variabilă este măsurată în prezent, iar alta este înregistrată din trecut, ideea de bază fiind aceea de a reflecta în ce măsură starea prezentă a putut fi anticipată de observaţiile colectate cu referire la trecut. În studiile transversale, toate variabilele sunt colectate în acelaşi timp, la momentul derulării studiului. În studiile prospective, o variabilă este măsurată în prezent, iar cel puţin o altă variabilă este măsurată ulterior, în viitor, încercându-se să se realizeze o legătură predictivă între cele două.

După cel de-al doilea criteriu, vorbim despre studii descriptive, despre studii predictive şi despre studii cauzal-explicative (cu valenţe exploratorii). Cele trei tipuri de studii au fost discutate pe larg în cel de-al treilea capitol al cărţii. Aici vom sublinia doar diferenţa dintre studiile predictive şi cele cauzal-explicative. Primele, deşi au în vedere studierea variabilelor psihologice într-o succesiune temporală, nu urmăresc identificarea unei relaţii de tip cauzal, ci doar surprinderea unor factori antecedenţi (de exemplu, realizarea de studii în care dificultăţile materiale din cadrul unei relaţii de cuplu prezic durata căsniciei).

Din combinarea celor două criterii de clasificare ar rezulta nouă tipuri diferite de studii non-experimentale. De exemplu, un studiu transversal descriptiv ar putea fi constituit de derularea unui studiu de screening pentru a cunoaşte care este prevalenţa cazurilor de depresie majoră diagnosticată clinic în rândul populaţiei de adolescenţi. În schimb, un studiu retrospectiv predictiv[2]

ar putea urmări măsura în care adolescenţii depresivi, faţă de un lot similar control, au fost expuşi în mai mare măsură în copilăria lor timpurie (până la 7 ani) la cazuri înregistrate oficial de violenţă domestică.

Deşi clasificarea respectivă poate fi una demnă de atenţie dintr-o perspectivă pedagogică, ea suferă de două mari neajunsuri. Primul, şi cel mai important dintre ele, se referă la faptul că un anumit tip de studii non-experimentale poate fi clasificat în categorii multiple. De exemplu, studiul de caz poate viza o abordare descriptivă transversală, cum la fel de bine poate fi centrat pe oferirea de posibile explicaţii (cu rol explorator) obţinute într-un demers longitudinal prospectiv. De asemenea, sondajele de opinie pot fi derulate cu rol descriptiv sau predictiv, vizând aspecte transversale sau retrospective ş.a.m.d. De aceea, clasificarea nu este una potrivită pentru a surprinde specificul anumitor tipuri particulare de studii non-experimentale. Cel de-al doilea neajuns se referă la titulatura aleasă pentru unul dintre criteriile de clasificare. Astfel, a spune despre studiile de tip non-experimental că pot urmări obiective explicative presupune o forţare a lucrurilor, pentru că aşa cum am descris pe parcursul acestei lucrări, studiile de tip non-experimental vizează mai degrabă obiective de tip exploratoriu, descriptiv sau descriptiv-analitic, din cauza multiplelor limite la nivelul posibilelor variabile confundate pe care le implică (absenţa controlului prin design). Din acest punct de vedere, o clasificare mai bună, ţinând cont de obiectivul studiului urmărit, este de a le împărţi în studii exploratorii, studii descriptive şi studii predictive (analitice), cele din urmă având o paletă largă de opţiuni care merg de la simpla asociere de variabile la situaţia inferării unor afirmaţii de tip cauzal (în cazul unor studii non-experimentale longitudinale).

Dată fiind această eterogenitate a poziţiilor de clasificare a tipurilor de studii, precum şi posibila suprapunere şi întindere a acestora pe multiple paliere, vom renunţa la includerea diferitelor studii non-experimentale în anumite categorii specifice, preferând o abordare pragmatică, bazată pe expunerea celor mai întâlnite tipuri de studii non-experimentale. De aceea, în

cele ce urmează vom prezenta pe rând patru categorii de studii transversale: studiile de caz, sondajele de teren, studiile corelaţionale şi studiile de diagnostic, un tip de studii retrospective – comparaţiile de tip *ex post facto*, cunoscute şi sub denumirea de studii caz-control (*case-control studies*) şi un tip de studii prospective (studiile longitudinale pe cohorte). Din motive de spaţiu tipografic, în cadrul fiecăruia dintre acestea vom sublinia doar elementele definitorii şi câteva aspecte aplicative importante. Fiecărui dintre aceste tipuri de studii i-ar putea fi dedicate capitole sau chiar cărţi întregi dacă ar fi abordate într-o manieră comprehensivă.

Înainte de a aborda fiecare dintre aceste tipuri majore de studii, vom prezenta în cadrul tabelului 5.1 principalele avantaje şi dezavantaje ale fiecăruia, pornind de la unele criterii menţionate de Petrie şi Sabin (2009) sau în WHO (2001).

Tabelul 5.1. *O sinteză a utilităţii şi limitelor principalelor tipuri de studii non-experimentale*

Aspect evaluat	Studiu de caz	Sondaj de opinie	Studiu corelaţional	Studiu de diagnostic	Comparaţie *ex post facto*	Studiu longitudinal
Aspectul descriptiv	Ilustrarea de cazuri prototipice sau a celora care diverg de la modelul teoretic	Gradul de extindere a problemei	Analiza asocierii dintre problema studiată şi alţi predictori (factori de risc sau protectori). Anumite designuri complexe, cum ar fi ecuaţiile structurale, pot testa criteriul	Gradul de corectitudine în predicţie	Poate fi util pentru a sugera relaţii de tip cauză-efect în cazul unor tulburări rare şi pentru a prezenta aspecte descriptive.	Evidenţa cazurilor noi
Aspectul predictiv/cauzal		Nu este potrivit pentru analiza cauzală		Aceste studii sunt foarte potrivite pentru obiective analitice de tip predictiv, însă nu şi pentru cele de tip cauzal.		Este cel mai potrivit design dintre cele non-experimentale pentru a fi utilizat în studii cu obiective cauzale. Potrivit pentru identificarea factorilor de

			statistic al unei relaţii cauzale.			risc şi a celor protectivi.
Aspect explicativ	Poate sugera explicaţii care vor fi ulterior testate adecvat	Nerecomandat pentru a oferi explicaţii pentru fenomenul observat	Nerecomandat pentru a oferi explicaţii pentru fenomenul observat	Nu vizează scopuri de tip cauzal-explicativ, fiind exclusiv studii predictive (cu rol prognostic)	Poate sugera explicaţii care vor fi ulterior testate adecvat	Poate sugera posibile explicaţii pentru fenomenul studiat, însă acestea au un caracter exploratoriu
Selecţia participanţilor	Cei afectaţi de problema studiată, îndeosebi cazurile rare	Participanţi neselecţionaţi (cu sau fără a fi afectaţi de problema studiată)	Participanţi neselecţionaţi (cu sau fără a fi afectaţi de problema studiată)	Se apelează la participanţi selecţionaţi special pe baza unor teste standard.	Participanţii sunt selecţionaţi astfel încât comparaţia să fie făcută între două loturi similare, cu excepţia variabilei criteriu utilizată pentru formarea grupelor.	Utilizarea de participanţi selecţionaţi sau neselecţionaţi este dependentă de tipul de design longitudinal implementat.
Impactul variabilelor confundate	Foarte ridicat din cauza riscului de subreprezentare a constructului	Redus, dacă există o definire clară a problemei cercetate	Destul de ridicat, din cauza problemei celei de-a treia variabile	Redus, deoarece sunt urmărite exclusiv scopuri prognostice sau de diagnostic, fără a infera	Destul de ridicat, fiind necesare măsuri adecvate de asigurare a similarităţii grupelor comparate	Destul de ridicat, îndeosebi legat de aspecte precum istoricul maturarea şi mai ales pierderea participanţilor

				aspecte cauzale		
Resurse implicate	Foarte reduse	Costuri financiare ridicate, timp redus de implementare	Destul de reduse, atât ca timp, cât şi financiar	Medii	Destul de reduse	Foarte mari, atât financiare, cât şi ca timp de implementare
Utilitate ridicată mai ales în studii ce urmăresc…	Generarea de ipoteze şi studii viitoare pornind de la cazuri particulare.	Stabilirea prevalenţei fenomenului.	Identificarea unor predictori relevanţi pentru fenomenul studiat.	Diagnosticul şi/sau screeningul unei probleme.	Sugerarea unor posibili factori cauzali sau antecedenţi (predictori), responsabili pentru diferenţele dintre loturi.	Stabilirea incidenţei fenomenului şi a unor relaţii de tip cauză-efect, acolo unde nu pot fi derulate studii experimentale.

5.1. Studiile de caz

Primul tip discutat de studii non-experimentale este studiul de caz. În esenţa sa, această strategie de cercetare vizează aspecte descriptive, ilustrative, cum ar fi acela de a reflecta printr-un demers deductiv o anumită teorie sau stare de fapt existentă la nivel general printr-unul sau mai multe cazuri particulare care să sprijine ideile cu un caracter mai ridicat de generalitate. Însă demersul logic poate fi la fel de bine unul inductiv, astfel încât pornind de la caracteristicile observate în situaţii particulare să fie emise anumite teorii sau explicaţii de natură psihologică, pentru o anumită stare de fapt. Din acest punct de vedere, Sigmund Freud rămâne cel mai cunoscut om de ştiinţă din sfera psihologiei şi care a conferit studiului de caz un rol central printre metodele de cercetare utilizate pentru elaborarea şi modificarea teoriilor sale psihanalitice.

De altfel, Kazdin (2010) oferă o foarte bună trecere în revistă a diferitelor roluri (scopuri) care pot fi trasate studiilor de caz. Astfel, un prim rol important al studiilor de caz este acela de a servi drept sursă de idei şi de ipoteze, de conceptualizare teoretică a unui fenomen, aspecte ce vor fi ulterior sprijinite sau nu prin dovezi empirice colectate din studii de alt gen. Această perspectivă conferă studiilor de caz dimensiunea exploratorie. Din acest punct de vedere, studiile de caz pot fi utile pentru a oferi insighturi nu numai despre aspectele teoretice implicate în cazul analizat, ci şi cu privire la eficienţa unor anumite tehnici utilizate în soluţionarea problemei respective. În al doilea rând, studiile de caz sunt mijloace excelente de descriere a unor fenomene rare. Studierea intensivă a acestor cazuri, care nu pot fi analizate prin alte metode de cercetare din motive pragmatice/de fezabilitate, poate oferi informaţii valoroase, îndeosebi de natură descriptivă, dar şi exploratorie. În al treilea rând, studiile de caz pot fi utilizate ca mijloc pentru a solicita revizuirea/regândirea unui anumit model teoretic, prin prezentarea acelor cazuri care contravin sau nu susţin anumite modele sau teorii existente în legătură cu fenomenul studiat. În sfârşit, fiind o abordare intensivă bazată pe o descriere detaliată a unei stări de fapt, studiile de caz au un rol pedagogic important, prin facilitarea înţelegerii nuanţelor ce însoţesc fenomenul studiat de către specialişti. Datorită acestor multiple roluri pe care le poate avea studiul de caz, Sechrest *et al.* (1996 *apud* Kazdin 2010) afirmau că acestea pot informa, intriga, inspira şi incita cititorii să aprofundeze problema studiată.

Din punct de vedere tehnic putem întâlni studii de caz singulare, respectiv serii de studii de caz. Cel din urmă aspect se referă la un set de studii de caz care vizează explorarea/descrierea aceleiaşi probleme analizate. De asemenea, obiectul studiului de caz este o unitate de analiză care poate fi de orice mărime, cum ar fi o persoană, o echipă de muncă, o organizaţie etc.

Formatul studiilor de caz nu este unul impus, însă cel mai frecvent conţine printre secţiuni: (a) o descriere a obiectivului studiului de caz/argumentarea selectării cazului respectiv; (b) date despre istoricul personal al individului

analizat sau despre contextul/specificul organizaţiei studiate; (c) descrierea problemei supuse analizei de caz; (d) o evaluare psihologică a persoanei investigate/o diagnoză a organizaţiei supusă analizei; (e) o discuţie a posibilelor cauze responsabile pentru starea de fapt respectivă; (f) intervenţii utilizate pentru ameliorarea/optimizarea situaţiei, precum şi (g) probleme întâmpinate şi/sau rezultate obţinute. Liantul dintre cele şapte secţiuni[3] specificate este oferit de perspectiva teoretică utilizată în studiul cazului respectiv. Astfel, abordarea teoretică are un rol important în realizarea studiilor de caz, fiind esenţială în realizarea de inferenţe pornind de la anumite manifestări comportamentele observabile, spre constructele psihologice abstracte cărora li se atribuie responsabilitatea pentru aceste manifestări comportamentale. O exemplificare a unui asemenea studiu de caz în domeniul clinic, din perspectiva teoriei raţional-emotive, este oferită de David (2006).

Caseta 5.1. *Specificul analizei datelor în studiile de caz*

Spre deosebire de celelalte tipuri majore de studii, cum ar fi studiile experimentale în care reglarea aspectelor de interes se realizează îndeosebi prin design, iar aparatul statistic este utilizat pentru a răspunde unor obiective de tip cauzal sau explicativ, în cazul studiilor de caz se apelează la o strategie diferită de analiză a datelor. Astfel, dat fiind caracterul explorator al multor studii de caz, modul de colectare şi analiză a datelor contribuie la formularea (sugerarea) unor mecanisme cauzale sau explicative prin apelul nemijlocit la capacitatea autorului de a analiza critic situaţia descrisă şi de a face inferenţe corecte. Hartley (2004) trece în revistă asemenea strategii, cum ar fi: (a) accentul pus pe surprinderea unor categorii (teme) centrale pentru problematica respectivă şi pe gradul de potrivire dintre un anumit tipar/model teoretic şi cazul prezentat (studii de caz în scop descriptiv); (b) referirea la date care sprijină o anumită explicaţie teoretică, în paralel cu respingerea unor explicaţii alternative (studii de caz exploratorii cu referire la potenţiale mecanisme cauzal-explicative). Din acest punct de vedere, în studiile de caz se procedează similar situaţiilor din alte strategii de cercetare, punându-se o valoare mai mare pe acele informaţii care permit inferenţe specifice, de sprijinire a unei anumite teorii/explicaţii şi care, totodată, elimină alte explicaţii alternative concurente.

Specificul studiilor de caz de a aprofunda înţelegerea unei probleme în contextul natural al individului/organizaţiei evaluate conduce însă şi la un

neajuns important, acela de a renunţa la controlul sistematic al variabilelor care ar putea pune în pericol validitatea acestor rezultate. De aceea, studiile de caz au un nivel scăzut al validităţii interne şi externe, aspect care poate fi compensat într-o anumită măsură prin strategia de a descrie o serie de cazuri care să „replice” ideile subliniate prin acest demers. În pofida acestor neajunsuri, studiul de caz reprezintă o strategie de cercetare respectabilă, fiind foarte utilă în scopuri exploratorii şi descriptive. O altă strategie de compensare a neajunsurilor unui studiu de caz, considerată a fi una superioară, este aceea de a apela la designul experimentelor realizate pe un singur individ (*single-case experimental design*). Detalii despre această strategie de cercetare pot fi găsite în Barlow, Nock şi Hersen, 2008, iar în limba română o scurtă introducere este disponibilă în Vîrgă, 2007.

5.2. Sondajele de opinie

Această strategie de cercetare, deosebit de cunoscută în spaţiul public românesc datorită apariţiei frecvente în mass-media a rezultatelor obţinute pe teme politice, are o denumire mai puţin fericită, deoarece ideea de a efectua un sondaj (*survey*) poate viza nu numai aspecte de opinie (de exemplu, „Ce marcă de bere preferaţi?”), ci şi aspecte factuale (de exemplu, „Aţi consumat bere în ultimele 24 de ore?”). Unii autori preferă utilizarea altor denumiri, precum anchetă de teren sau anchetă sociologică[4]. Din nefericire, nici aceste nume nu sunt cele mai potrivite pentru a descrie întreaga sferă de cuprindere a sondajelor, dacă luăm în considerare progresele tehnologice evidente (multe anchete nu se mai derulează pe teren, ci prin intermediul internetului sau telefonic), respectiv aria de aplicabilitate (deşi sondajul este un instrument de bază în domeniul sociologiei, aplicabilitatea sa nu se rezumă doar la acest domeniu al cunoaşterii).

Din anumite puncte de vedere, sondajele reprezintă opusul studiilor de caz. Dacă studiile de caz vizează surprinderea aspectelor unice, a excepţionalului, focalizându-se asupra fenomenului într-o manieră intensivă, dar nesistematică, sondajele de opinie vizează surprinderea aspectelor generale, a tendinţelor comune, focalizându-se într-o manieră extensivă şi sistematică asupra fenomenului studiat. De asemenea, dacă studiile de caz urmăresc predominant o metodologie calitativă de colectare şi interpretare a rezultatelor, sondajele sunt, prin definiţie, o strategie de cercetare cantitativă din prisma colectării, analizei şi interpretării rezultatelor. Însă ambele au drept element comun faptul că pot servi unor scopuri de cercetare de tip descriptiv sau exploratoriu.

Prin modalitatea riguroasă, sistematică, de planificare, derulare şi analiză a rezultatelor, sondajul poate fi evaluat mai uşor din perspectiva validităţii concluziilor sale comparativ cu situaţia unui studiu de caz. De aceea, în paginile următoare vom detalia o serie de puncte importante în planificarea şi derularea unui sondaj de opinie, adresând problema eşantionării, a volumului eşantionului, a ratei de răspuns, dar şi aspecte legate de construirea chestionarului şi de selectare a celei mai bune strategii de colectare a datelor. Ne vom rezuma doar la cazul celor implementate transversal (un singur val de măsurare), problematica sondajelor derulate într-un context longitudinal fiind abordată în secţiunea următoare (5.3), dedicată studiilor pe cohorte.

5.2.1. Problematica eşantionării

Sondajele de opinie presupun chestionarea unei anumite părţi din populaţia de interes, cu scopul de a infera aspecte cu un nivel ridicat de generalizare (la nivelul populaţiei din care a fost extras eşantionul). Astfel, spre deosebire de situaţia unui referendum, în care cel puţin la nivel teoretic este inclusă în analiză întreaga populaţie investigată, în cazul sondajelor sunt extraşi doar anumiţi participanţi pentru a fi chestionaţi. În limba română, unele şcoli de sociologie şi psihologie fac diferenţe terminologice între un eşantion de

participanţi şi un lot de participanţi, în funcţie de modul în care a avut loc selecţia. Termenul „eşantion” este rezervat situaţiei în care extragerea din cadrul populaţiei de interes s-a făcut prin metode probabilistice de selectare a participanţilor, iar „lot” este utilizat în cazul în care selecţia s-a făcut prin metode neprobabilistice. Diferenţa dintre cele două categorii de metode este importantă, deoarece doar metodele probabilistice (aleatorii) de selectare a participanţilor asigură şanse egale de a fi extraşi tuturor indivizilor din cadrul unei populaţii, o caracteristică ce minimizează riscul de a obţine rezultate distorsionate. Cel mai simplu exemplu pentru a înţelege diferenţele dintre cele două forme este reprezentat de sondajele realizate de diferite posturi TV în timpul emisiunilor. Să luăm de exemplu, postul de ştiri Antena 3. În timpul unei emisiuni politice, moderatorul adresează telespectatorilor întrebarea: „Susţineţi candidatura lui Crin Antonescu la preşedinţia României?”, aşteptând răspunsurile pe forumul şi pe pagina de Facebook a emisiunii. Rezultatele indică faptul că 90% dintre telespectatori exprimă o opinie favorabilă faţă de această candidatură, iar concluzia moderatorului este aceea că există o susţinere clară din partea românilor faţă de acest actor politic. Dar este rezultatul unul valid? Este foarte probabil ca procentul obţinut de către Crin Antonescu să fie unul distorsionat faţă de realitate, în sensul că el nu a fost obţinut pe baza unui eşantion probabilistic, ci în baza unuia neprobabilistic. De ce se întâmplă acest lucru? Pentru că studiul s-a bazat (a) doar pe opinia celor care sunt telespectatori ai Antenei 3 şi ai emisiunii respective (cei care nu au vizionat emisiunea şi/sau care nu urmăresc postul respectiv de televiziune nu au avut şanse egale de a fi cuprinşi în eşantion) şi (b) chiar dintre cei care au urmărit emisiune, o bună parte, îndeosebi cei din mediul rural, nu au putut răspunde dat fiind că răspunsurile necesitau acces la internet. Să ne imaginăm care ar fi fost rezultatul dacă sondajul era iniţiat de un post TV cu viziuni politice diferite, precum B1 TV.

Prin urmare, între metodele probabilistice sunt cuprinse doar acelea care oferă şanse egale de a fi selectaţi în eşantion indivizilor din populaţia de

interes, în timp ce metodele neprobabilistice nu oferă şanse egale, selectarea subiecţilor în studii făcându-se pe baza altor criterii, cum ar fi cel al oportunităţii (disponibilităţii) de a fi incluşi în studiu. În caseta 5.2 sunt trecute principalele tehnici de eşantionare utilizate în mod curent, cu precizarea că prin populaţie nu se are în vedere ceea ce se înţelege în termeni comuni (de exemplu, populaţia României), ci populaţia variabilei de interes (de exemplu, populaţia firmelor mici şi mijlocii, cu regim juridic de tip microîntreprindere; populaţia persoanelor divorţate; populaţia studenţilor la psihologie din centrul universitar Timişoara etc.).

Caseta 5.2. *Tehnici de eşantionare probabilistice şi neprobabilistice*

Tehnici probabilistice

Tehnica de eşantionare aleatorie simplă

Presupune selectarea din cadrul unei liste alfabetice ce conţine populaţia de interes (de exemplu, masteranzi din Universitatea de Vest din Timişoara) a unui număr de participanţi egal cu volumul stipulat al eşantionului. De exemplu, dacă se doreşte extragerea unui număr de 30 de studenţi din listele respective ce cuprind un total de 3.000 de participanţi posibili, se poate apela la tabele cu numere aleatorii sau la aplicaţii informatice cu extragere aleatorie a 30 de numere de la 1 la 3.000, fără posibilitatea repetării numărului deja extras (în limbaj informatic este o procedură *quasi-random*). Tehnica este eficientă (a) dacă populaţia nu este una foarte mare, deoarece pentru populaţii mari, situate în diverse locaţii, costurile de acest tip cresc, sau (b) în contextul în care aplicarea are loc în mediul electronic, fiind însă dependentă de existenţa unei baze de date cu toate cazurile eligibile (adrese de e-mail) spre a fi incluse în eşantionare.

Tehnica de eşantionare aleatorie sistematică

Reprezintă o derivare de la metoda aleatorie simplă, apelându-se la un pas de eşantionare. De exemplu, într-un studiu în care se doreşte selectarea probabilistică a 100 de participanţi dintr-o populaţie de 1.000 de locuitori (raportul de selectare fiind 1 din 10), se poate alege într-o manieră aleatorie un număr de la 1 la 10 (de exemplu, numărul 7), iar apoi să fie selectaţi participanţii pe baza acestui pas de 10 (de exemplu, persoanele având numerele 7, 17, 27, … 997). Avantajul acestei tehnici faţă de cea aleatorie simplă este acela că poate fi implementată şi în absenţa unei liste nominale. De exemplu, poate fi chestionată fiecare a zecea gospodărie dintr-un cartier, atâta timp cât se cunoaşte cu aproximaţie volumul populaţiei de investigat pentru a calibra pasul de eşantionare într-un mod adecvat (de exemplu, dacă populaţia este constituită din 200 de unităţi şi se doreşte un eşantion de 40 de unităţi se poate utiliza un pas de 5).

Tehnica de eşantionare cluster

Tehnica de eşantionare de tip cluster presupune aplicarea principiului probabilistic (egalitate de şanse) asupra categoriei (grupului) din care face parte individul vizat. De exemplu, într-un studiu efectuat pe studenţi de la nivelul ciclului de licenţă de la toate specializările, ar fi dificil din punct

de vedere practic de utilizat un algoritm de selectare precum cel al evaluării aleatorii sistematice sau al celei simple. În schimb, ar fi mai uşor să se selecteze aleatoriu câteva grupe de studenţi (de exemplu, grupa a II-a din anul II de la specializarea Finanţe-Bănci; grupa I de la specializarea Psihologie din anul III etc.), urmând a fi chestionaţi toţi studenţii din grupele respective. Avantajul acestei tehnici este de natură practică, fiind mult mai uşor de colectat datele printr-un asemenea demers. Această facilitate vine însă cu un cost, eşantionarea de tip cluster fiind metoda cu cel mai ridicat risc de distorsiune a rezultatelor dintre metodele de tip probabilistic.

Tehnica de eşantionare stratificată

Această tehnică este utilă în cazul în care se doreşte a se ţine seama de anumite caracteristici importante ale populaţiei de interes (de exemplu, într-un studiu realizat pe studenţi ar fi important să luăm în considerare anul de studii în care se află aceştia). Astfel, se poate apela la o strategie aleatorie simplă sau la una sistematică, care să nu fie aplicată pe întreaga listă de studenţi de la nivelul de studii licenţă, ci separat, pe fiecare listă de studenţi, cea a studenţilor din anul I, a celor din anul II, respectiv pe a celor din anul III. O asemenea strategie are sens atunci când asemenea criterii (precum anul de studii), denumite în termeni sociologici straturi (*strata*), pot avea un impact asupra tipului de răspuns relevant şi dorim să controlăm acest lucru. Din punct de vedere statistic, eşantionarea stratificată prezintă cel mai scăzut risc de rezultate distorsionate dintre metodele probabilistice amintite până acum, desigur, dacă straturile utilizate au o relevanţă mare asupra tipului de răspuns oferit. De asemenea, acest tip de eşantionare este util atunci când se doresc a fi realizate analize la nivel de categorii/subgrupe5. Din punct de vedere practic este mai dificil de implementat o asemenea eşantionare. Ea necesită resurse de timp şi costuri suplimentare, mai ales dacă se urmăresc criterii combinate, cum ar fi surprinderea concomitentă a anului de studii şi a statutului studentului (bugetar sau cu taxă), în observarea răspunsurilor colectate. Desigur, în interiorul fiecărui strat sau combinaţii de straturi, selecţia participanţilor se face după criterii probabilistice, precum cele amintite anterior.

Tehnica de eşantionare multifazică (multistadială)

Tehnica respectivă reprezintă, de fapt, combinarea secvenţială a tehnicilor probabilistice de tip cluster cu alte tehnici menţionate anterior. Spre ilustrare, dacă Ministrul Educaţiei doreşte să implementeze un sondaj de opinie destinat elevilor de liceu din anul terminal pentru a afla opiniile acestora faţă de ideea introducerii bacalaureatului profesional, se poate recurge la următoarea schemă de eşantionare multifazică cu trei stadii succesive: (a) stratificată – gruparea unităţilor şcolare în funcţie de două criterii: poziţionarea liceului (în reşedinţa de judeţ *vs* în altă localitate din judeţ), respectiv în funcţie de performanţa anterioară la bacalaureat (licee cu procent de promovabilitate în anul anterior de peste 75%, licee cu promovabilitate cuprinsă între 25% şi 75%, respectiv licee cu o promovabilitate sub 25%); (b)

eşantionare de tip cluster (de exemplu, din lista cu licee cu o promovabilitate medie situate în afara reşedinţei de judeţ sunt selectate 5% dintre unităţile şcolare – fiecare al 20-lea liceu din lista respectivă); (c) tot eşantionare de tip cluster (chestionarea tuturor elevilor din cadrul unei clase a XII-a din unităţile selecţionate, cu precizarea că dacă există mai multe clase terminale în liceul selectat, să fie chestionată clasa care se situează cel mai aproape de intrarea principală din şcoală).

Ce ar însemna acest lucru? Să presupunem că 5% din unităţile şcolare înseamnă un număr de 50 de unităţi (dintr-un total de 1.000 de licee), iar din fiecare unitate este selectată o clasă (cu
o mărime medie de 25 elevi per clasă). S-ar totaliza un număr de 1.250 de elevi de an terminal chestionaţi, repartizaţi procentual în funcţie de două criterii considerate a fi importante – performanţa liceului şi mediul de provenienţă al acestuia.

Eşantionarea multistadială (multifazică) este utilizată îndeosebi în contexte în care populaţia este alcătuită dintr-un număr mare de participanţi, cu o compoziţie eterogenă şi în care este dificil să se implementeze o eşantionare de tip aleatoriu dintr-o singură secvenţă (operaţie).

Tehnici neprobabilistice

Eşantionarea pe bază de convenienţă

Destul de mulţi studenţi confundă acest tip de eşantionare cu una probabilistică, metoda fiind de departe cea mai întâlnită formă de eşantionare din domeniul psihologiei. Apelarea la cunoştinţe sau la colegi pentru a completa un set de chestionare, utilizarea persoanelor întâlnite pe stradă, cererea opiniilor telespectatorilor etc. constituie o eşantionare pe bază de convenienţă, deoarece nu se oferă şanse egale tuturor respondenţilor potenţiali. De exemplu, într-un studiu în care un student apelează la colegii săi pentru a completa un chestionar, lotul constituit nu poate fi considerat un eşantion extras din populaţie, deoarece (a) este improbabil ca populaţia de interes să fie aceea a studenţilor cu statut de cunoştinţe, ci studenţii în general; (b) chiar dacă populaţia cu statut de cunoştinţe ar fi cea de interes, nu au fost oferite şanse egale nici măcar acestora, deoarece au fost testate doar cunoştinţele prezente/disponibile la momentul evaluării.

Ca urmare a acestui neajuns, rezultatele obţinute nu pot fi generalizate direct la nivelul întregii populaţii de interes, prin urmare validitatea externă a studiului are de suferit. Pentru a compensa această stare de fapt, strategia alternativă de creştere a validităţii externe are în vedere replicarea rezultatelor în studii diferite, chiar dacă acestea s-au derulat pe loturi obţinute prin eşantionare neprobabilistică pe bază de convenienţă.

Eşantionarea ţintită

Acest tip de eşantionare (*purposive sampling*), deşi este tot una neprobabilistică, poate fi utilă în contexte în care se investighează populaţii specifice, greu identificabile. De exemplu, în studii despre somnambulism sau derulate pe persoane cu o orientare homosexuală, ar fi dificil şi nefezabil să ne bazăm pe o eşantionare generală din care să extragem doar acele cazuri de interes. O variantă derivată din eşantionarea ţintită este cea de tip avalanşă (*snowball sampling*). În acest caz fiecare persoană chestionată este rugată să ofere referinţe/date de contact despre alte persoane care ar corespunde profilului căutat, având astfel acces la o

adevărată reţea socială. Deşi acest tip de eşantionare poate fi util în identificarea indivizilor cu anumite caracteristici de interes întâlnite rar, metoda prezintă aceleaşi dezavantaje pe care le au şi celelalte metode neprobabilistice de eşantionare.

Eşantionarea pe cote

Eşantionarea pe cote îşi propune să elimine o parte din neajunsurile tehnicilor neprobabilistice, printr-o serie de măsuri menite să inducă impresia de reprezentativitate. De exemplu, într-un studiu realizat pe o populaţie de studenţi la psihologie, în care există statistici care arată faptul că 80% dintre aceştia sunt de sex feminin, respectiv că 80% urmează cursurile unor universităţi particulare, se poate urmări respectarea acestor proporţii în lotul de respondenţi studiat. De exemplu, dacă lotul conţine 200 de studenţi, acesta ar trebui format din 40 de bărbaţi (8 de la universităţi de stat şi 32 de la instituţii private) şi din 160 de femei (32 de la universităţi de stat şi 128 de la cele private), astfel încât se respectă proporţia de gen şi de tip de instituţie urmată, de la nivelul populaţiei de interes. Deşi acest tip de eşantionare pare a fi unul onorabil, nici el nu respectă principiul egalităţii de şanse în selecţie, persoanele fiind selectate în cotele respective, tot pe bază de convenienţă. De aceea, în lipsa utilizării unor criterii aleatorii de selecţie, deşi au fost prestabilite anumite cote în cazul unor criterii relevante, există în continuare un risc crescut de a obţine rezultate distorsionate prin absenţa controlului asupra altor posibile variabile relevante. De exemplu, rezultatele ar putea fi diferite dacă majoritatea răspunsurilor colectate de la studenţii unor universităţi de stat provin din centre universitare clasate în topul performanţei sau de la universităţi cu specializări mai puţin performante în psihologie.

Din perspectiva analizei validităţii rezultatelor unor sondaje, utilizarea unor metode probabilistice de eşantionare este singura cale care conduce la creşterea validităţii externe a rezultatelor prin posibilitatea de a generaliza (extinde) rezultatele obţinute de la nivelul eşantionului studiat la nivelul populaţiei de interes din care a fost extras eşantionul. Datele provenite din sondaje realizate fără asigurarea unei metode aleatorii de selectare a respondenţilor prezintă un nivel ridicat de risc la adresa validităţii externe a rezultatelor, însă ele pot fi utile în scopuri descriptive sau exploratorii, din perspectiva altor tipuri de validitate precum este cea statistică sau cea de construct.

5.2.2. Problema mărimii eşantionului

Investiţia de efort făcută cu ocazia selectării unui eşantion prin metode probabilistice nu este răsplătită decât pe jumătate în cazul în care volumul eşantionului nu este suficient de mare pentru a reduce amplitudinea intervalului de încredere asociat rezultatului obţinut la un nivel acceptabil. Astfel, la baza reţetei succesului unui sondaj bine efectuat se regăseşte asocierea dintre utilizarea unei tehnici probabilistice pentru a selecta indivizii cuprinşi în eşantion şi asigurarea unui număr suficient de participanţi pentru a obţine reprezentativitatea rezultatului la nivelul întregii populaţii de interes din care a fost extras eşantionul[6].

Pentru a determina mărimea unui eşantion se are în vedere: (a) tipul de metodă probabilistică de eşantionare utilizată; (b) nivelul de eroare prespecificat; (c) nivelul de încredere în rezultate (probabilitatea ca ele să fie adevărate); (d) frecvenţa anticipată a scorului care constituie punctul central de interes. Primul dintre cele patru elemente determină tipul de formulă utilizată, celelalte constituindu-se ca puncte variabile între care poate jongla cercetătorul pentru a atinge un maximum de eficienţă (acurateţe) cu un minimum posibil de costuri.

Formula pentru stabilirea numărului de subiecţi necesari a fi incluşi într-un studiu, pentru o eşantionare probabilistică simplă şi o populaţie numeroasă (de cel puţin 10 ori mai mare decât volumul estimat al eşantionului) ar fi conform lui Crano şi Brewer (2002) următoarea:

$$n = \frac{p(1-p)}{SE^2}$$

unde p este procentul estimat oferit pentru un anumit răspuns, iar SE este eroarea standard, adică mărimea erorii pe care o putem tolera.

De exemplu, într-un sondaj pe teme politice, suntem angajaţi pentru a determina procentul estimat de voturi pe care l-ar putea înregistra o alianţă politică în condiţiile în care sondajele anterioare indicau că va obţine un

procent de 25%. În acest caz, p ia valoarea 0,25 (restul, de 1 – p = 0,75, fiind votanţii altor formaţiuni politice). De asemenea, ne dorim să avem o eroare de maxim ±0,04 (4%), ceea ce înseamnă aproximativ 2 × SE, pentru un interval de încredere de 95%, adică un SE de 0,02. Aplicând formula mai sus menţionată, estimăm că vom avea nevoie de aproximativ 469 de persoane (0,25 × 0,75/0,02 × 0,02), pentru a afla cu o eroare de ±4%, la un interval de încredere de 95% procentul real pe care l-ar obţine alianţa politică analizată. În schimb, într-un alt studiu în care se doreşte a se investiga starea de fapt asupra unei probleme despre care nu există niciun fel de date statistice anterioare, se porneşte de la o valoare p egală cu 0,50 (50% şanse ca opinia să fie una favorabilă sau una nefavorabilă). În acest caz, aplicând aceeaşi formulă vom avea nevoie de 625 de participanţi pentru o eroare de ±4% la un interval de încredere de 95% [0,50 × (1 – 0,50)/(0,02 × 0,02)] sau chiar de 2.500 de persoane, pentru a asigura o eroare de doar ±2% la acelaşi interval de încredere de 95% [0,50 × (1 – 0,50)/(0,01 × 0,01)]. Cu excepţia unor situaţii justificate, când preferinţele sunt foarte apropiate şi se doreşte o diferenţiere fină între acestea, opţiunea aleasă va lua în considerare cel mai bun raport cost-beneficiu. Pentru situaţia descrisă, am opta probabil pentru un eşantion de 625 de persoane în defavoarea celui format din 2.500 de persoane (costurile fiind de aproximativ 4 ori mai mici, în contextul în care eroarea de eşantionare creşte puţin, de la ±2% la ±4%).

Mici ajustări ale acestei formule, cu un impact redus la nivelul volumului eşantionului, pot fi introduse: (a) dacă se are în vedere o altă metodă probabilistică decât cea simplă aleatorie; (b) dacă sondajul vizează o populaţie finită (în care raportul dintre totalul populaţiei şi volumul necesar al eşantionului este de cel mult 10); (c) dacă se ţine cont de procentul real de respondenţi efectivi (de exemplu, din 100.000 de votanţi eligibili, participă doar 70% la vot). Detalii despre aceste ajustări sunt disponibile în Ahmed (2009) sau Crano şi Brewer (2002). Fără a intra în detalii tehnice aici, se cuvine să precizăm că, la acelaşi număr de indivizi cuprinşi într-un eşantion,

cea mai precisă estimare se obţine printr-o eşantionare aleatorie stratificată, urmată de o eşantionare aleatorie simplă, în timp ce un nivel ceva mai scăzut de precizie se obţine prin eşantionarea de tip cluster, mai ales dacă se apelează la investigarea unui număr redus de clustere de mari dimensiuni (10 departamente a câte 50 de indivizi fiecare) decât dacă se apelează la un număr mare de clustere de dimensiuni reduse (50 de departamente, a câte 10 indivizi fiecare).

Desigur, aceste calcule merită făcute în cazul utilizării unor metode probabilistice. În cazul apelului la metode neprobabilistice, formulele amintite nu îşi au sensul, volumul lotului fiind relevant doar din perspectiva puterii statistice de testare a ipotezelor. În cazul utilizării de metode de eşantionare neprobabilistice există un risc crescut de rezultate distorsionate faţă de situaţia existentă la nivelul populaţiei. O excepţie de la această stare de fapt este cazul în care între lotul extras şi populaţia din care acesta face parte raportul este apropiat de o valoare unitară.

5.2.3. Rata de răspuns vs rata refuzurilor

Chiar dacă s-a apelat la o metodă de eşantionare aleatorie şi la un număr potrivit de participanţi, din punct de vedere teoretic rămâne un risc ridicat de distorsiune a rezultatelor dacă se constată o rată ridicată a refuzurilor de a răspunde. De exemplu, dacă doar 20% dintre cei selectaţi iniţial în eşantion au acceptat să răspundă, iar 80% nu, se poate ajunge la o distorsiune în rezultate (de exemplu, s-ar putea obţine un procent mai mare de răspunsuri extreme, fie ele favorabile sau nefavorabile, decât s-ar fi obţinut dacă toţi cei selectaţi iniţial ar fi răspuns).

Rata de răspuns se calculează făcând raportul simplu dintre cei care au răspuns şi totalul celor contactaţi pentru a oferi răspunsuri. De exemplu, dacă doar 250 de persoane au răspuns la un sondaj în care au fost contactaţi 1.000 de participanţi eligibili, vom spune că există o rată de răspuns de 25% (sau o

rată de refuz de 75%). Deşi teoretic sunt preferabile sondajele cu o rată cât mai ridicată de răspuns, pentru a reduce riscul apariţiei unei distorsiuni de selecţie, în practică, au fost sesizate diferenţe minore în acurateţea datelor între studiile cu rată scăzută şi cele cu rată ridicată de participare (Holbrook, Krosnick şi Pfent, 2007). Prin urmare, această problemă este mai puţin severă comparativ cu celelalte două aspecte discutate: extragerea eşantionului prin metode aleatorii, respectiv stabilirea numărului optim de subiecţi.

5.2.4. Construirea chestionarului

Un sondaj bine realizat nu implică doar aspecte tehnice legate de selectarea şi construirea eşantionului, ci acordă la fel de multă atenţie aspectelor calitative legate de construirea şi formatul întrebărilor. Întregul efort depus pentru a apela la metode de eşantionare probabilistice poate fi năruit dacă lipsesc abilităţile de a dezvolta eficient întrebările vizate. De exemplu, o întrebare formulată ambiguu, de genul „Practicaţi sport regulat?” (răspunsuri: „Da” sau „Nu”), poate induce rezultate distorsionate pentru că unii pot răspunde afirmativ înţelegând prin termenul „regulat” ideea practicării săptămânale a unui sport, în timp ce alţii, care fac la fel de mult sport, pot răspunde negativ dacă înţeleg prin aceeaşi sintagmă ideea de sport practicat zilnic. Tabelul 5.2 prezintă o listă de probleme în construirea întrebărilor/a chestionarului în ansamblul său, precum şi soluţiile aferente, pornind de la o sinteză prezentată în Neuman (1997).

Tabelul 5.2. *Sugestii pentru o redactare adecvată a întrebărilor dintr-un chestionar*[7]

Problema identificată	*Formulare inadecvată*	*Formulare adecvată*
Aspecte legate de imprecizia formulării		
evitarea abrevierilor, a jargonului, a	ONG complexul lui Oedip	Organizaţii neguvernamentale O definiţie funcţională care să însoţească termenul tehnic.

unor termeni tehnici care pot crea dificultăţi de înţelegere		
evitarea ambiguităţii	Practicaţi sport în mod regulat?	Să se înlocuiască termenul „regulat” cu un aspect cuantificabil. Practicaţi un sport: (a) zilnic; (b) de câteva ori pe săptămână; (c) săptămânal etc.
evitarea dublei chestionări	Vă plac filmele SF şi cele de groază?	Se construiesc două întrebări separate, una care să vizeze filmele SF, cealaltă filmele de groază.
evitarea întrebărilor cu răspuns dificil	Câţi litri de benzină consumaţi în medie pe an?	Câţi litri de benzină consumaţi în medie pe săptămână?
evitarea negaţiilor şi a dublei negaţii	Absolvenţilor de liceu nu ar trebui să li se permită să se înscrie la facultate dacă nu au promovat bacalaureatul?	O exprimare directă care să nu îngreuneze înţelegerea textului: Absolvenţii de liceu se pot înscrie la facultate dacă promovează examenul de bacalaureat.
lipsa de ancore în oferirea răspunsului, dacă este nevoie de o estimare	În medie, cât de mult alcool consumaţi într-o lună?	Segmentarea întrebării în contexte concrete (în baruri, acasă; în vizite la prieteni/rude), urmată de o subsumare a acestora.
Aspecte legate de inducerea (sugerarea) unui răspuns		
evitarea exprimărilor cu încărcătură emoţională ridicată sau care pun presiune suplimentară pe respondent	Politica organizaţiei este de a arăta o grijă suplimentară faţă de problemele de mediu, prin reducerea cantităţii de documente tipărite. Susţineţi această poziţie?	Susţineţi măsura de a reduce cantitatea de documente tipărite?
evitarea sugestiilor din întrebare	Ar trebui primăria să acorde mai mulţi bani pentru asfaltarea drumurilor?	Selectaţi care ar fi cel mai important domeniu care ar trebui finanţat de primărie: (a) problemele de canalizare; (b) asfaltarea drumurilor etc.
evitarea falselor premise	Pentru a reduce lungimea rândului de aşteptare, aţi fi de acord cu extinderea programului de lucru a biroului de relaţii cu publicul?	Cum apreciaţi situaţia orarului de lucru a biroului de relaţii cu publicul: (a) ar putea fi extins (b) ar putea fi păstrat aşa; (c) ar putea fi redus.

evitarea chestionării situaţiilor ipotetice (intenţiilor)	Să presupunem că s-ar deschide un nou magazin ABC pe strada ta. Ai cumpăra de acolo?	Preferabilă ar fi chestionarea comportamentelor curente (de exemplu, unde îşi face în prezent cumpărăturile, ca tip de magazin (mall *vs* de cartier), ca frecvenţă (zilnic *vs* săptămânal) şi ca localizare (distanţa faţă de casă).
reducerea răspunsurilor cu un nivel ridicat al dezirabilităţii sociale	Subiectele sensibile conduc la o subestimare a activităţilor negative (de exemplu, conducerea maşinii fără a purta centura de siguranţă) şi la o supraestimare a frecvenţei activităţilor pozitive (de exemplu, mersul la bibliotecă).	Strategii multiple precum: (a) garanţii explicite pentru păstrarea confidenţialităţii; (b) sublinierea importanţei colectării unei situaţii reale; (c) reformularea întrebărilor la persoana a III-a; (d) apelul la sondaje derulate fără contact vizual (cum ar fi cele derulate prin internet).
diminuarea efectului de ordine (de reportare)	Se aplică două întrebări cu o ordine fixă: atitudinea faţă de conducerea maşinii sub influenţa alcoolului; apoi faţă de viol.	Alternarea ordinii celor două întrebări sau utilizarea unei strategii care presupune trecerea dintre întrebări generale (abstracte) spre cele specifice.
diminuarea efectului capului de listă	Citirea unei liste cu un număr destul de ridicat de răspunsuri în aceeaşi ordine (de exemplu, alfabetică).	Rotirea (contrabalansarea) răspunsurilor sau confecţionarea de cartonaşe cu fiecare răspuns, astfel încât ordinea de citire a lor de la un respondent la altul să fie una aleatorie.
Alte aspecte care pot distorsiona rezultatele		
lipsa unor categorii adecvate de răspuns	Lucraţi sau sunteţi şomer? Vârsta Dvs.: (a) între 14 şi 20; (b) între 20 şi 25 de ani; (c) între 25 şi 30 de ani.	Nu sunt luate în seamă toate categoriile posibile (studenţii, pensionarii etc.). Vârsta în ani împliniţi: (a) între 14 şi 20; (b) între 21 şi 25; (c) între 26 şi 30 de ani.
utilizarea de întrebări-filtru	Variante care forţează oferirea unui răspuns. „Cât de mult vă place berea X?”: (a) foarte mult; (b) mult; (c) puţin; (d) foarte puţin.	Fie sunt utilizate mai multe variante de răspuns, inclusiv NŞ (nu ştiu/nu o cunosc), fie se introduce mai întâi o variabilă filtru („Aveţi o părere despre berea X?”).
grijă faţă de distorsiunile posibile în cazul întrebărilor de cunoştinţe	Apreciaţi activitatea politică a lui Nicolae Dică?	Surprinzător, un număr destul de însemnat de respondenţi ar răspunde la întrebarea respectivă, deşi personajul respectiv a fost un fotbalist. Ar putea fi utilizate întrebări de control (capcană) de genul celei prezentate, pentru a diferenţia între cei care fac aprecieri în cunoştinţă de cauză şi celelalte cazuri.

5.2.5. Selectarea celei mai potrivite strategii de aplicare a chestionarului

Sondajele pot fi implementate fie direct, pe teren (prin operatori care realizează interviuri faţă în faţă cu persoanele din eşantionul selectat) sau telefonic (fără a beneficia de o relaţie faţă în faţă cu interlocutorul), fie indirect, printr-o abordare clasică (chestionare trimise prin poştă) sau una modernă (prin intermediul internetului şi/sau al sms-urilor). Decizia de a opta pentru una sau alta din variantele menţionate este dependentă de doi factori: (a) costurile implicate; (b) caracteristicile sondajului. Tabelul 5.3 sintetizează câteva dintre cele mai importante criterii de care ar trebui să se ţină seama în alegerea modalităţii de implementare a sondajului. De exemplu, dacă sondajul este unul scurt şi pe o temă cunoscută publicului larg, iar datele colectate trebuie prelucrate şi analizate cât mai rapid, s-ar putea apela la sondajul telefonic, deoarece acesta corespunde cel mai bine celor trei criterii menţionate. În mod similar, dacă populaţia de interes este una selecţionată (de exemplu, masteranzi), iar subiectul abordat este unul delicat (aspecte factuale şi de opinie despre plagiat şi copiat la examene), se poate apela la un sondaj derulat pe internet, dacă există acces la informaţii personale ale studenţilor cum ar fi adresa de email a acestora pentru a putea păstra ideea extragerii aleatorii a listei de respondenţi, cu posibilitatea păstrării anonimatului (printr-o legătură către pagina web unde se găseşte chestionarul).

Tabelul 5.3. *Direcţii posibile de implementare în funcţie de caracteristicile sondajului de opinie*

Aspect vizat	*Sondaj pe teren*	*Sondaj telefonic*	*Sondaj prin poştă*	*Sondaj pe internet*
Buget necesar	ridicat	mediu	redus	redus
Lungimea chestionarului	mare	mică	medie	medie
Durata perioadei de colectare	medie	scurtă	foarte lungă	lungă
Rata răspunsurilor	mare	medie	mică	mică
Teme sensibile	inadecvat	parţial	adecvat	adecvat

		adecvat		
Nivelul de educaţie al respondentului	irelevant	puţin relevant	studii medii şi superioare	studii medii şi superioare
Accesul la întreaga populaţie	ridicat	relativ ridicat	ridicat	mediu (urban)/scăzut (rural)
Explicitarea aspectelor complexe (neclarităţilor)	posibilă	redusă	absentă	absentă

5.2.6. Alte aspecte relevante în construirea şi derularea sondajelor de opinie

În cazul sondajelor de opinie realizate faţă în faţă sau telefonic se apelează la operatori de teren. Dincolo de aspectele tehnice deja menţionate, corectitudinea rezultatelor este dependentă de calitatea instruirii acestora, de motivarea lor de a colecta precis datele şi de măsurile de control postaplicare ale calităţii datelor colectate. Acest lucru este important de cunoscut pentru a reduce tendinţa operatorilor de a fabrica datele şi/sau de a nu respecta procedura de eşantionare convenită.

Dacă problema calităţii operatorilor afectează sondajele directe, îndeosebi studiile de teren, problematica acurateţei listelor din care se extrage eşantionul de participanţi prin metode probabilistice afectează toate tipurile de sondaje, îndeosebi pe cel realizat prin internet, secondat de sondajul telefonic. De exemplu, în SUA, deşi 95% dintre gospodării au un telefon, doar 75% dintre acestea sunt identificabile public, fiind trecute în cartea de telefon. Acest lucru înseamnă că nu există şanse egale pentru toţi participanţii de a fi selectaţi în studiu, deoarece unii dintre ei nu se regăsesc pe lista din care se extrage în mod aleatoriu eşantionul studiat (Crano şi Brewer, 2002). Pentru a suplini asemenea deficienţe au fost concepute o serie de strategii alternative de lucru, cum ar fi, în cazul sondajelor telefonice, apelarea aleatorie (*random digit dialing*) (de exemplu, se extrage un număr din cartea de telefon, iar ultimele două cifre sunt regenerate automat pentru a suna aleatoriu la un nou număr de

telefon, ce poate fi inclus sau nu în cartea respectivă de telefon). Mesajul legat de această problemă a listelor (a bazei de date existente) este acela că nivelul calitativ al implementării procedurii de randomizare este dependent de calitatea listei ce conţine întreaga populaţie de interes (în exemplul oferit, 5% din populaţie, care nu deţine un telefon, fiind totuşi exclusă din cercetare). În cazul sondajelor de teren, problema poate fi una similară, în sensul că unele persoane se pot afla în situaţia de a nu se regăsi pe listele concepute de autorităţi. Să ne amintim de întreaga discuţie privind acurateţea listelor de electori cu ocazia ultimului referendum de suspendare a preşedintelui Traian Băsescu. Pentru a reduce problema inadvertenţelor din liste, în cazul sondajelor de teren se obişnuieşte a se apela la o rută de deplasare care să asigure şanse egale de a accesa toate gospodăriile de interes, nu doar pe acelea înregistrate corect în listele oficiale.

În sfârşit, o atenţie deosebită trebuie acordată nuanţelor în formularea întrebărilor din chestionare, pentru a viza direct aspectul urmărit şi pentru a nu se produce confuzii, cum ar fi aceea de a măsura opinii în locul unor aspecte factuale, comportamentale. De exemplu, se folosesc exprimări diferite pentru a urmări comportamente („Aţi băut bere în ultimele 24 de ore?”), obiceiuri („Cât de frecvent aţi băut bere în ultima lună?”), intenţii („Aţi cumpăra bere pentru a servi la masă în weekenduri?”), aspecte învăţate („Care este primul lucru care vă vine în minte când vă gândiţi la bere?”), preferinţe („Care este berea care vă place cel mai mult?”) sau atitudini („Care mărci de bere corespund afirmaţiei «este o bere de calitate»?”) (Datculescu, 2006).

În această secţiune ne-am rezumat la prezentarea sondajelor de opinie aplicate transversal. Deşi o asemenea strategie este una utilă pentru a descrie starea de fapt existentă la un moment dat, ea suferă de aceeaşi problemă pe care o au şi studiile corelaţionale şi, în general, orice design de cercetare transversal. Metoda poate evidenţia gradul de răspândire a unui fenomen (prevalenţa), cât şi eventuale contingenţe existente între anumite variabile, însă nu poate surprinde relaţii de tip cauză-efect. Motivul este unul simplu, acela că

pentru a surprinde relaţii cauzale una dintre condiţii este aceea de a observa că modificările variabilei-cauză le precedă temporal pe cele ale variabilei-efect, iar acest lucru nu poate fi demonstrat într-un studiu cu o singură testare. Pentru a adresa asemenea întrebări într-un context non-experimental, se pot derula sondaje într-un design longitudinal, cum sunt studiile pe cohorte sau cele panel.

5.3. Studiile longitudinale pe cohorte

Înainte de a prezenta mai multe detalii despre studiile pe cohorte, se cuvine să facem o serie de precizări terminologice relevante pentru a înţelege avantajele şi dezavantajele studiilor longitudinale în general, nu doar a celor efectuate pe cohorte. Apoi vom pune accent pe prezentarea specificului studiului longitudinal pe cohorte, utilizat cu precădere în sfera clinică şi care se apropie foarte mult de zona studiilor panel, utilizate în ştiinţele sociale. În sfârşit, în cea de-a treia parte a acestei secţiuni vom prezenta o serie de informaţii practice, utile pentru implementarea şi analiza unor asemenea studii.

5.3.1. Tipologia studiilor longitudinale şi obiectivele care pot fi adresate

Toate tipurile de studii longitudinale se aseamănă prin faptul că cercetătorul culege datele de la participanţii la studiu în mai multe etape diferite, denumite valuri (*waves*). Diferenţele dintre tipurile de studii longitudinale (reflectate în figura 5.1) sunt discutate în tabelul 5.4.

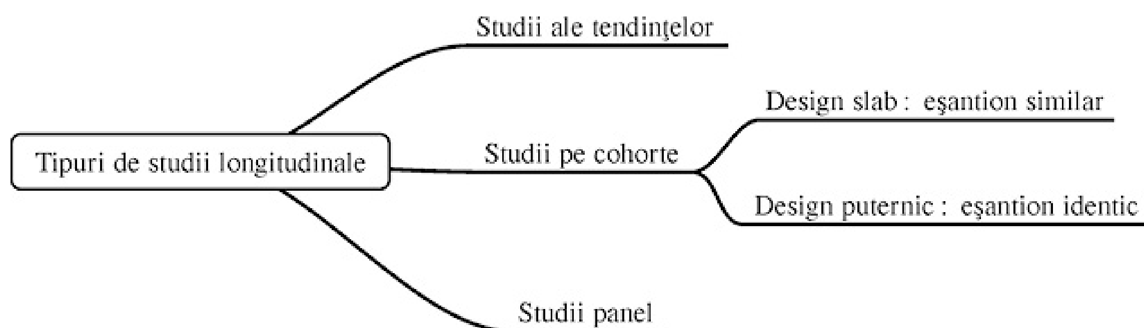


Figura 5.1. *Tipuri majore de studii care apelează la un design longitudinal*

Tabelul 5.4. *Ilustrarea principalelor tipuri de studii longitudinale*

Tipul de studiu	**Descrierea caracteristicilor de bază**	**Ilustrare**
Studiul tendinţelor (*trend study)*	Aceste studii sunt utile pentru a surprinde direcţia de evoluţie a unui domeniu prin testări repetate. Specificul lor constă în faptul că nu sunt testaţi repetat aceiaşi respondenţi, ci în fiecare val de testare se extrage un eşantion aleatoriu din populaţia de interes.	Pentru a surprinde evoluţia apelului populaţiei la servicii de psihoterapie, se poate urmări anual, timp de zece ani, evoluţia procentului persoanelor care apelează la servicii de psihoterapie. În fiecare an se extrage câte un eşantion reprezentativ din populaţia generală şi se stabileşte prevalenţa utilizării acestui serviciu. Nu sunt testate aceleaşi persoane, decât din întâmplare. Ideea este de a urmări tendinţa de evoluţie a serviciilor de psihoterapie (creşte, stagnează sau scade frecvenţa utilizării lor în populaţie), prin apelul la eşantioane reprezentative extrase din populaţia respectivă, fapt care permite compararea datelor colectate anual.
Studiul pe cohortă (*cohort study*)	Sub această denumire identică există două tipuri diferite de studii majore de cohortă. Ele se aseamănă prin faptul că ambele apelează la un tip special de eşantion denumit cohortă (participanţi care împărtăşesc una sau mai multe caracteristici comune, de exemplu orfani de mamă înainte de vârsta de 5 ani, crescuţi apoi în familii monoparentale), dar se diferenţiază prin strategia de selectare a participanţilor.	În versiunea cohortelor cu testarea repetată a unor eşantioane similare, dar nu identice (tradiţia din sociologie), pentru a surprinde evoluţia utilizării serviciilor de psihoterapie la nivelul cohortelor de interes, ambele cohorte sunt urmărite anual, timp de zece ani, pentru a evalua frecvenţa apelului la servicii de psihoterapie, precum şi diferenţele existente între cele două cohorte (generaţii). De exemplu, în primul val sunt selectaţi aleatoriu câte 200 de participanţi născuţi în

A. Într-un caz, bazat pe tradiţia cercetărilor sociologice, *sunt utilizaţi participanţi diferiţi* de la un val de testare la altul, dar *aparţinând aceleiaşi cohorte* (de exemplu, un eşantion reprezentativ de persoane născute în perioada de glorie a comunismului [1976-1980] este testat în valul 1, în timp ce în valul 2, care are loc zece ani mai târziu, este selectat un alt eşantion reprezentativ al aceleiaşi cohorte). În studiile realizate pe cohorte sunt selectate de obicei cel puţin două sau mai multe cohorte diferite, pentru a realiza comparaţii în evoluţia observată între cei aparţinând unei cohorte (de exemplu, născuţi între 1976 şi 1980) şi cei aparţinând unei alte cohorte (de exemplu, născuţi în perioada 1991-1995).

B. În cel de-al doilea caz, având drept sursă sfera medicinei, în mod special a epidemiologiei, sunt utilizaţi *aceiaşi participanţi aparţinând unei cohorte*, aceştia fiind testaţi în mod repetat, în două sau mai multe valuri[9].

De exemplu, dacă se compară cohorta celor născuţi în perioada 1976-1980, cu cea a celor născuţi în perioada 1991-1995, sunt colectate

perioada 1976-1980, respectiv în perioada 1991-1995, în baza unei eşantionări aleatorii a populaţiei generale cu domiciliul în Timişoara. Anul următor sunt extraşi alţi 200 de indivizi născuţi în perioada 1976-1980, respectiv alţi 200 de participanţi născuţi în perioada 1991-1995. În mod similar se procedează în fiecare an din cei zece supuşi perioadei de urmărire, cu precizarea că nu se are în vedere testarea aceloraşi subiecţi, ci testarea unor indivizi care aparţin cohortei de interes, având în vedere faptul că extragerea eşantionului se face într-un mod aleatoriu, pentru a asigura posibilitatea de a generaliza rezultatele la nivelul cohortei din care fac parte şi a face comparaţii intergeneraţionale.

În versiunea cohortelor cu testarea repetată a unor eşantioane identice (tradiţia din medicină) pentru surprinderea evoluţiei serviciilor de psihoterapie şi a diferenţelor dintre cohorte, sunt testaţi repetat timp de zece ani, aceiaşi 400 de indivizi (200 aparţinând cohortei 1976-1980 şi 200 din cohorta 1991-1995). Desigur că şi în cazul acestora, pentru a asigura generalizarea rezultatelor la nivelul populaţiilor de interes, selecţia iniţială se face prin proceduri probabilistice de eşantionare.

date de la aceiaşi subiecţi, cele două eşantioane fiind testate repetat pe perioada de urmărire.

În subtipul cel mai răspândit de studii pe cohortă în tradiţia medicală există o condiţie suplimentară, aceea ca toţi cei 400 de indivizi selectaţi să nu fi avut la momentul iniţial (T0) nicio experienţă anterioară în care să fi apelat la servicii psihoterapeutice. Cu alte cuvinte, cele două cohorte comparate să fie similare, cu excepţia

		criteriului de diferenţiere, pentru a vedea în ce măsură generaţia din care fac parte (dinainte sau de după revoluţia din 1989) joacă rolul unui factor facilitator sau inhibitor în apelul la servicii de psihoterapie.
Studiul panel (*panel study*)	Studiile panel presupun testarea repetată (în două sau mai multe valuri) a aceluiaşi eşantion de persoane selectat iniţial, de preferinţă, tot prin proceduri aleatorii de eşantionare[8]. Deşi ideea studiilor panel se apropie foarte mult de ideea studiilor pe cohorte din tradiţia epidemiologiei, există unele diferenţe între ele. De exemplu, în studiile panel accentul nu este pus pe compararea cohortelor, ci pe a urmări evoluţia în timp a aceloraşi indivizi incluşi în studiu. De asemenea, ideea unor criterii de eligibilitate, prezentă atât de frecvent în studiile de cohortă (de exemplu, să nu fi apelat anterior la servicii de psihoterapie), nu are o însemnătate la fel de mare în studiile panel.	Pentru a surprinde evoluţia apelului populaţiei la servicii de psihoterapie se poate urmări anual, timp de zece ani, evoluţia procentului persoanelor care apelează la servicii de psihoterapie. Cu ocazia primului val (primul an) se extrage un eşantion reprezentativ din populaţia generală şi se stabileşte prevalenţa utilizării acestui serviciu. Apoi este urmărită evoluţia aceloraşi subiecţi timp de zece ani, astfel încât fiecare persoană inclusă în studiu este testată anual, în zece valuri succesive.

Diferenţele dintre cele patru forme majore de studii longitudinale[10] sunt importante din perspectiva obiectivelor urmărite. Astfel, deşi studierea tendinţelor este cel mai uşor de implementat (pentru că nu trebuie urmărite aceleaşi persoane şi nici nu trebuie avută o grijă deosebită în selectarea participanţilor după anumite criterii), ea nu permite adresarea unor întrebări de tip cauză-efect, deoarece nu se urmăreşte evoluţia în timp a aceloraşi persoane, ci mai degrabă o tendinţă generală de evoluţie a lucrurilor. Mai mult, uneori această tendinţă generală poate ascunde rezultate interesante care ar putea fi

observate doar printr-o abordare diferenţiată, de genul studiilor de cohortă (vezi exemplul ipotetic descris în figura 5.2).

De exemplu, în cazul unui studiu al tendinţelor privind consumul de droguri uşoare precum canabisul, se constată o relativă stagnare a prevalenţei consumului în jurul unei medii de 6,75% consumatori din cadrul populaţiei generale investigate. Această stagnare poate ascunde însă tendinţe diferite (unele de creştere, altele de descreştere, anulându-se reciproc), dacă ne raportăm la aceleaşi rezultate (o medie generală de 6,75%) obţinute pe trei cohorte diferite (trei generaţii care aveau vârste medii de 10, 20, respectiv 30 de ani la data demarării studiului în 1991). Astfel, se observă că spre deosebire de cei care fac parte din cohorta 1960-1962, la care se observă o prevalenţă a consumului de canabis în jurul a 7%, generaţiile 1970-1972, respectiv 1980-1982 ating în medie o prevalenţă de peste 9% la aceeaşi vârstă de 30 de ani, semnal care ar putea fi interpretat ca o creştere a consumului de droguri la generaţiile mai noi. Acest lucru este reflectat şi prin compararea consumului de droguri la 20 de ani între cohorta din anii 1970-1972 (în jurul a 6% din respondenţi) şi cohorta din anii 1980-1982 (în jurul a 9% dintre respondenţi). De asemenea, se constată o descreştere a consumului de canabis după împlinirea vârstei de 30 de ani la ambele cohorte la care s-au colectat date după împlinirea acestei vârste. Toate aceste nuanţe nu ar fi putut fi surprinse printr-o simplă analiză globală a tendinţei, ele fiind puse în lumină numai printr-o abordare diferenţială, de comparare a evoluţiei diferitelor cohorte evaluate.

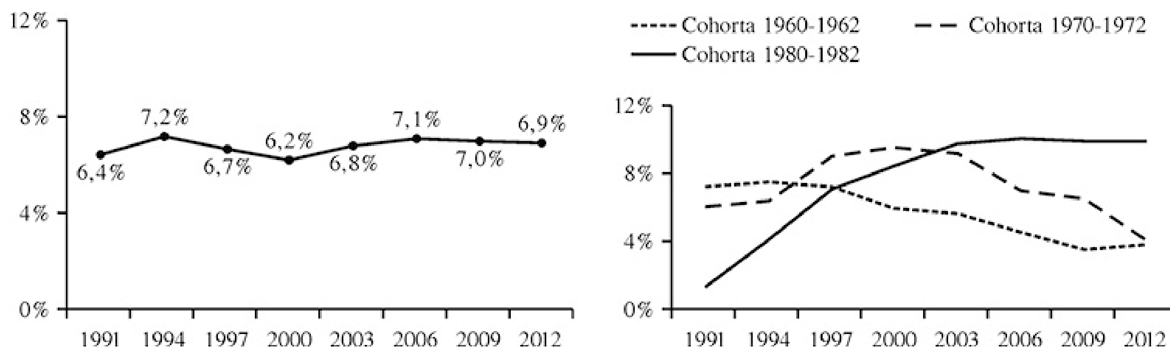


Figura 5.2. *Rezultate ipotetice ale unui studiu despre consumul de droguri: studiu al tendinţelor (partea stângă)* vs *studiu pe trei cohorte generaţionale (partea dreaptă)*

Informaţii suplimentare despre specificul fiecărui tip de studii longitudinale pot rezulta şi din analiza obiectivelor ce pot fi studiate prin apelul la studii longitudinale (vezi tabelul 5.5).

Tabelul 5.5. *Relaţia dintre obiectivele studiului şi tipurile de studii longitudinale*

Obiective	**Descriere (tip de studiu adecvat)**
1. Detectarea modificărilor în timp şi patternul acestora.	1.A. Identificarea schimbărilor şi a formei acestora (liniară, exponenţială etc.). Ilustrare: *Se modifică satisfacţia studenţilor de la UVT faţă de programele educaţionale oferite? (T)* 1.B. Identificarea diferenţiată a schimbărilor (schimbarea este aceeaşi pentru toţi indivizii sau există schimbări diferite în funcţie de anumite caracteristici ale studenţilor?). Ilustrare: *Se modifică diferit satisfacţia studenţilor din cohorta celor finanţaţi de la bugetul de stat faţă de cea a studenţilor plătitori de taxă? (C1)* 1.C. Analiza relaţiilor dintre schimbările observate (dacă anumite schimbări observate sunt asociate cu alte modificări – schimbări reciproce). Ilustrare: *Schimbările în atitudinea adolescenţilor faţă de consumul de alcool sunt asociate cu modificări în atitudinea lor faţă de fumat? (T)* 1.D. Analiza diferenţială a relaţiilor dintre schimbările observate (dacă există diferenţe interindividuale în intensitatea asocierii anumitor schimbări). Ilustrare: *În cazul fumătorilor, modificarea atitudinii faţă de consumul de tutun este asociată cu modificarea atitudinii faţă de consumul de cafea, în timp ce în cazul nefumătorilor nu există o asociere între magnitudinea modificării uneia dintre ele şi schimbarea celeilalte. (C1)* 1.E. Identificarea numărului de cazuri noi (incidenţa unei probleme). Ilustrare: *Care este incidenţa anuală a consumului de droguri la adolescenţi? (C2 sau P)*
2. Indicarea direcţiei şi a magnitudinii unei relaţii de tip cauzal.	2.A. Identificarea unor variabile-cauză (a unor determinanţi/factori antecedenţi) responsabile pentru o anumită schimbare apărută (de ce se schimbă indivizii?). Ilustrare: Burnout-*ul este un factor favorizant al absenteismului. (C2 sau P)* 2.B. Analiza diferenţială a impactului variabil al factorilor antecedenţi asupra comportamentului sau atitudinii urmărite (de ce efectul variabilei cauză asupra efectului este diferit la anumiţi indivizi faţă de alţii?). Ilustrarea: *O analiză a rolului moderator al schimbării organizaţionale (angajaţi din companii cu acţionariat nou* vs *angajaţi din companii cu acelaşi acţionariat) în relaţia dintre* burnout *ca un prezumtiv factor favorizant şi absenteism. (C2)*

Legendă: T – studiu al tendinţelor; C1 – studiu pe cohorte cu eşantioane similare; C2 – studiu pe cohorte cu eşantioane identice; P – studiu panel

Tabelul 5.5 poate fi privit în sistem piramidal. De exemplu, studiile de tip panel pot fi oricând utilizate în locul celor de tendinţă pentru a surprinde obiectivele mai puţin pretenţioase. De asemenea, studiile pe cohorte efectuate pe eşantioane identice (C2) pot înlocui studiile de acelaşi gen bazate pe eşantioane similare (C1). Însă aspectele 1A-1D ale primului obiectiv pot fi adresate şi prin studii longitudinale uşor de implementat, deoarece nu apelează la testarea aceloraşi participanţi (studiile T şi C1). Aceste studii se apropie mai mult de ideea studiilor transversale, nefiind vorba de testări repetate realizate pe aceeaşi participanţi. De aceea, ele sunt utile pentru a sesiza patternuri de asociere (patternuri ale modificărilor în timp ale variabilelor de interes), chiar dacă nu permit inferarea unor afirmaţii privind relaţiile cauză-efect.

5.3.2. Studiile pe cohorte (cu eşantion identic)

Aşa cum rezultă din tabelul 5.5 şi din explicaţiile însoţitoare acestuia, acest tip de studii poate fi utilizat, în principiu, pentru a răspunde la oricare dintre întrebările tipice întâlnite într-un studiu longitudinal, cu precizarea că este foarte util mai ales în situaţiile în care se urmăresc obiective de tip cauzal-descriptiv. De asemenea, acest tip de studii este cel mai întâlnit tip de studii longitudinale în domeniul clinic (medical), fiind cel mai utilizat design longitudinal din sfera psihologiei clinice, cea mai răspândită ramură aplicativă a psihologiei. Din aceste motive, am ales să prezentăm în secţiunea de faţă mai multe detalii caracteristice despre această strategie de cercetare.

Prin cohortă se înţelege un grup de persoane care împărtăşesc aceeaşi caracteristică într-o anumită perioadă de timp (cum ar fi cei născuţi în acelaşi an, cei care au absolvit un program masteral în străinătate, cei care provin dintr-o anumită etnie, fumătorii, primii născuţi într-o familie etc.). Cohorta poate fi comparată cu situaţia generală din populaţia din care a fost extrasă sau

cu un alt grup de control comparativ (de exemplu, cohorta fumătorilor este comparată cu un grup similar sociodemografic, dar compus din nefumători). În practică este aproape imposibil ca toate variabilele potenţial confundate să fie controlate într-un design non-experimental, dar cu toate acestea se obişnuieşte să se colecteze, în afara variabilei principale de interes, date despre evoluţia altor variabile potenţial confundate, pentru a controla prin mijloace statistice impactul acestora asupra criteriului urmărit. De exemplu, într-un studiu efectuat de Kempainnen *et al*. (2001) privind efectul ordinii născuţilor (primii născuţi, ultimii născuţi, cei singuri la părinţi, altă situaţie) asupra apariţiei schizofreniei, s-au măsurat şi controlat aspecte precum: (a) complicaţii perinatale prezente sau absente; (b) sarcină dorită *vs* nedorită; (c) numărul de fraţi şi surori; (d) tipul de familie la vârsta de 14 ani (monoparentală *vs* normală).

Figura 5.3 ilustrează exemplul tipic de design longitudinal întâlnit în studiile pe cohorte prospective, într-un studiu care urmăreşte să vadă măsura în care consumul excesiv de alcool este un antecedent al infracţionalităţii. De exemplu, din cadrul unei populaţii generale de adolescenţi se extrage un eşantion probabilistic cu un număr mare de participanţi (de exemplu, N = 3.000), care îndeplinesc următoarele două condiţii de eligibilitate pentru a intra în studiu: (a) au vârsta de 15 ani împlinită, respectiv (b) nu au în istoricul lor personal nicio problemă cu legea înregistrată în evidenţele poliţiei. Tuturor celor care îndeplinesc aceste două condiţii de eligibilitate li se aplică un test de screening privind dependenţa de alcool din care este selectat un grup de persoane care prezintă un risc crescut de dependenţă de alcool (percentilul 90 sau peste) (N = 300) şi un grup de control similar ca volum, extras randomizat din restul de 90% de participanţi selectaţi iniţial, dar având scoruri percentile sub 90 la testul de screening privind dependenţa de alcool. Apoi, cele două loturi selecţionate special sunt urmărite timp de zece ani, fiind înregistrate măsurători obiective ale încălcărilor legii (existenţa unor contravenţii,

respectiv a unor infracţiuni) după 5, respectiv după 10 ani de la începutul derulării studiului.

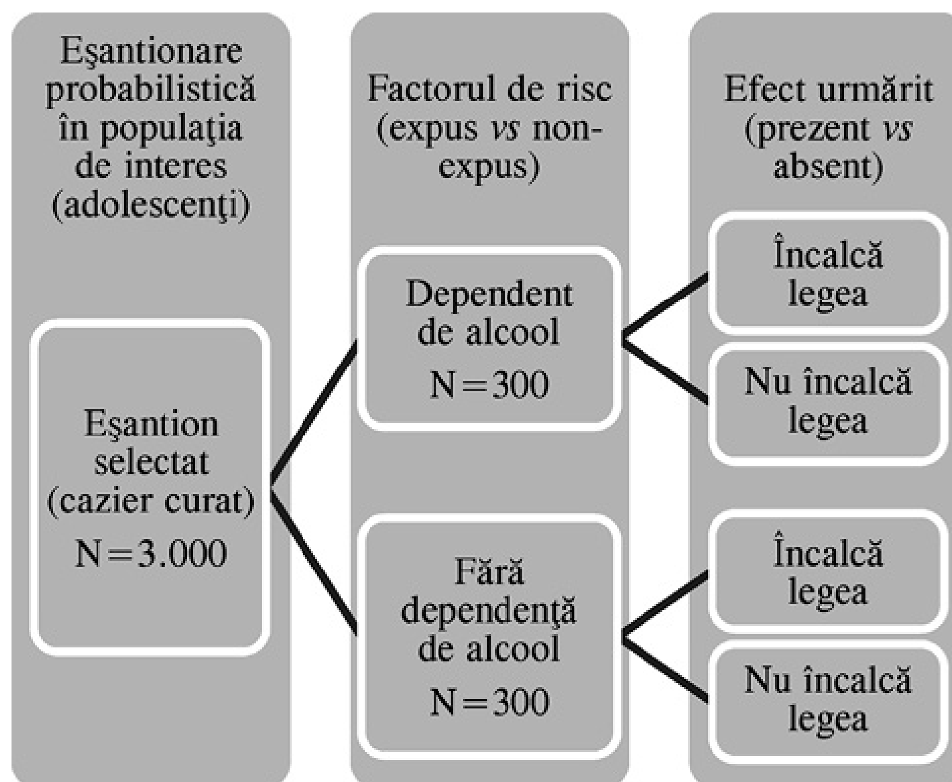


Figura 5.3. *Schema logică a unui studiu de cohortă prospectiv*

Datele colectate în cadrul unui studiu de cohortă nu se referă numai la prezumtivul factor cauzal (de risc) şi la efectul urmărit, ci este obligatoriu să se aibă în vedere colectarea datelor cu referire la unele variabile confundate. Dacă ar fi să facem referire la exemplul ilustrat în figura 5.3, ar fi prematur să afirmăm că dependenţa de alcool ar contribui la creşterea infracţionalităţii (în contextul în care s-ar constata că adolescenţii identificaţi ca fiind dependenţi de alcool cu ocazia screeningului comit mai multe infracţiuni decât adolescenţii din grupul de control). În schimb, ar fi valid să afirmăm că alcoolul este un predictor al stării infracţionale viitoare. Pentru a încerca să emitem inferenţe de tip cauzal, ar fi bine să măsurăm, pe lângă problematica alcoolului, nivelul altor variabile posibil confundate[11] precum: (a) stilul parental – cunoscându-se, de exemplu, o asociere între consumul excesiv de

alcool şi un stil parental *laissez-faire*; (b) consumul de droguri, asociat adesea cu consumul excesiv de alcool; (c) nivelul de educaţie; (d) statutul social economic; (e) tendinţele antisociale măsurate printr-o scală din zona psihopatologiei etc.)[12].

Dacă după ajustarea rezultatelor, ţinând cont de efectul acestor variabile confundate, vom găsi în continuare că alcoolul reprezintă un factor de risc pentru săvârşirea de infracţiuni, vom putea fi ceva mai încrezători că dependenţa de alcool ar putea fi o posibilă cauză pentru infracţionalitate. În schimb, dacă după ajustarea rezultatelor, prin luarea în considerare a impactului variabilelor potenţial confundate măsurate, consumul excesiv de alcool nu mai are un impact semnificativ statistic asupra comportamentului infracţional al tinerilor, vom putea afirma, cu o probabilitate destul de mare, că alcoolul nu reprezintă un factor cauzal, ci, asemenea săvârşirii de infracţiuni, consumul diferenţiat de alcool este doar un efect al unei alte variabile confundate. De exemplu, din analiza datelor ar fi posibil ca un stil parental *laissez-faire* să fie un antecedent cauzal atât pentru dependenţa de substanţe (inclusiv alcool), cât şi pentru apariţia unor conduite din zona infracţionalităţii.

Mai multe detalii despre situaţia în care putem emite afirmaţii de tip cauzal în baza unui studiu longitudinal de tip cohortă se regăsesc în tabelul de mai jos.

Tabelul 5.6. *Ghidul Bradford-Hill cu privire la interpretarea într-o manieră cauzală a rezultatelor din studiile longitudinale pe cohorte (adaptare după Bruce, Pope şi Stanistreet, 2008)*

Criteriu cauzal (necesar)	**Explicare şi ilustrare**
Factorul potenţial (expus) precedă efectul[13]	Principiul de bază în orice relaţie de tip cauză-efect este acela că factorul-cauză precedă temporal apariţia efectului. De exemplu, de la momentul expunerii la radiaţii este nevoie de câţiva ani până la apariţia leucemiei. Studiile pe cohortă pot compara un lot de persoane sănătoase, dintre care unii au fost expuşi la radiaţii (din zona Cernobîl) cu un lot similar, dar neexpus la radiaţii.
Intensitatea asocierii	Pe măsură ce intensitatea asocierii este una mai puternică, cu atât mai mare este probabilitatea de a fi în faţa unei relaţii de tip cauzal, deoarece oricâte ajustări ar fi determinate de factori potenţial confundaţi, tot nu ar putea

	elimina asocierea observată. Un exemplu este relaţia puternică dintre fumat şi cancer, în care riscul relativ are o valoare de aproximativ 10 (fumătorii prezintă un risc de 10 ori mai mare de a face cancer decât nefumătorii).
Menţinerea relaţiei independent de efectul variabilelor confundate	Este important să se demonstreze faptul că relaţia rămâne în picioare, chiar dacă am controlat statistic influenţa altor factori relevanţi măsuraţi. Din păcate, la acest capitol, rămâne dezavantajul major al studiilor non-experimentale, spre deosebire de cele experimentale, de a nu putea controla toate sursele cauzale posibile.
Consistenţă la nivel de doză (expunere diferită)	De exemplu, într-un studiu care urmăreşte impactul evenimentelor cu potenţial emoţional traumatic asupra apariţiei anxietăţii generalizate, acest criteriu este îndeplinit dacă se observă o incidenţă mai mare a anxietăţii generalizate la cei cu un istoric cu multiple stări traumatice (cum ar fi pierderea unui părinte; abuzul fizic; observarea unei situaţii de criză – incendierea casei etc.), decât în cazul celor care au fost expuşi doar la un singur eveniment cu potenţial traumatic major.
Consistenţă din studii şi populaţii diferite	Nivelul de încredere în ideea unei relaţii de tip cauză-efect provenită din studii non-experimentale longitudinale creşte dacă rezultatul este concordant cu cele provenite din alte studii.
Plauzibilitate	Ar fi destul de riscant să se afirme despre o asociere identificată că ar avea o natură cauzală dacă nu există un suport empiric şi teoretic pentru aceasta. În exemplul oferit, ar fi destul de greu de argumentat că excesul de alcool ar fi cauza directă a conduitei infracţionale viitoare, deşi s-ar putea argumenta că reprezintă un factor cauzal distal (consumul excesiv de alcool duce la destructurarea capacităţii de adaptare la cerinţele societăţii, ceea ce poate deschide calea comportamentelor infracţionale, ca modalitate de coping la inadaptarea socială şi la neajunsurile financiare).

Avantajul major al studiilor longitudinale efectuate pe acelaşi eşantion, cum sunt studiile pe cohortă cu design puternic şi studiile panel, este acela de a putea decela mai bine care dintre variabile ar putea fi variabile-cauză şi care sunt variabile-efect. De exemplu, în studiile corelaţionale transversale în care se observă o relaţie negativă între depresie şi stimă de sine este imposibil de stabilit dacă în urma intensificării simptomatologiei depresive apar modificări în direcţia scăderii stimei de sine sau dacă persoanele cu stimă de sine scăzută sunt mai vulnerabile/predispuse să ajungă la manifestări depresive. În mod similar, într-un studiu transversal care indică o asociere între expunerea la

material video violent şi comportamentul agresiv, nu se poate stabili clar dacă: (a) persoanele agresive preferă mai degrabă să urmărească material ce conţine scene violente sau (b) dacă acestea activează un nivel crescut de agresivitate la cei care vizionează materialul respectiv sau (c) dacă între cele două aspecte există o condiţionare reciprocă. Prin studiile longitudinale ce implică măsurători efectuate pe aceeaşi indivizi se poate stabili clar care dintre variabilele analizate apare mai întâi. Un alt avantaj important al studiilor pe cohorte este posibilitatea de a evidenţia impactul pe care îl are un anumit factor de risc (de exemplu, fumatul) asupra unor efecte multiple (cancer, boli cardiovasculare, greutatea corporală, rezistenţa fizică etc.), cu menţiunea că autorii acestor studii trebuie să adopte un comportament etic (cum ar fi evitarea tendinţei de a raporta rezultatele într-un mod trunchiat, precum doar cele semnificative, din multitudinea de date colectate) (Grimes şi Schulz, 2002).

Cu toate aceste avantaje şi plusuri sesizate în cazul studiilor pe cohorte, întrebarea naturală ar fi „de ce sunt ele mai rar întâlnite în domeniul psihologiei?". Principalul motiv este unul de ordin economic. Realizarea unui asemenea studiu necesită resurse uriaşe atât din perspectiva numărului de participanţi incluşi în studiu (cu care trebuie să se păstreze legătura pentru a fi evaluaţi periodic şi pentru a reduce pierderea participanţilor[14]), cât şi din punct de vedere al resursei de timp (în special în cazul studiilor prospective, în care trebuie să treacă un număr de ani de la momentul T0 până la momentul finalizării studiului). La acestea se adaugă complexitatea şi durata evaluărilor efectuate, fiind măsurate pe lângă variabilele de interes (posibila relaţie cauzală urmărită), multiple variabile potenţial confundate care trebuie măsurate pentru a fi controlate statistic. Astfel, urmărirea a 1.000 de persoane timp de 30 de ani, pentru a sesiza valoarea predictivă pe care stilul de ataşament format în copilăria timpurie îl are asupra comportamentului ulterior al individului în relaţiile sale romantice, nu este un demers uşor de realizat, iar

fondurile disponibile pentru acest gen de studii sunt mult mai dificil de găsit comparativ cu situaţia existentă în alte domenii, cum este cel al medicinei.

Caseta 5.3. *Alte detalii de luat în seamă în planificarea unui studiu pe cohorte (Grimes şi Schulz, 2002; WHO, 2001)*

Criterii clare pentru ce înseamnă grupul expus **vs** ***neexpus, precum şi pentru efectele urmărite (distorsiune în selecţie şi în măsurare)***

De exemplu, într-un studiu în care s-ar urmări efectul consumului de tutun asupra evoluţiei IQ-ului (degradării cognitive) este foarte important să se definească ce înseamnă să consume tutun. În cazul în care se formează mai multe grupe (consum accentuat, consum moderat, consum absent) este important să se definească foarte clar ce înseamnă acest lucru. De exemplu, în cazul consumului accentuat ar putea fi pus drept criteriu de includere ca participanţii să fumeze de cel puţin 6 luni de zile, cel puţin 20 de ţigări pe zi, pornind de la un standard de 1 mg de gudron, 0,1 mg de nicotină şi 2 mg de monoxid de carbon. Apoi trebuie avute criterii clare (inclusiv la nivelul extinderii volumului cohortelor pentru a menţine o putere statistică acceptabilă) pentru cazurile de încrucişare – indivizi iniţial nefumători care devin fumători înrăiţi sau persoane cuprinse iniţial în grupul fumătorilor înrăiţi care se lasă de fumat ulterior.

În mod similar ar trebui operaţionalizat clar ce se înţelege prin degradarea cognitivă (de exemplu, scăderi în nivelul IQ dincolo de valoarea unui index al modificării reale – vezi tabelul 4.7 din capitolul 4).

Dincolo de pragurile valorice stabilite este importantă şi metoda de colectare a datelor, cele mai bune fiind socotite metodele obiective, iar cele mai puţin dorite fiind cele bazate pe autoraportare. Problema încrederii în datele obţinute se pune mai ales în cazul studiilor pe cohorte retrospective, deoarece este destul de greu de apreciat corectitudinea datelor obţinute din trecut.

Selectarea adecvată a unui grup de control (distorsiune în selecţie)

Ideal ar fi ca grupul de control selectat să fie unul care să manifeste caracteristici similare cu cohorta de interes. De exemplu, într-un studiu ipotetic care să compare efectul ordinii de naştere dintr-o familie (primii născuţi *vs* ultimii născuţi) asupra apariţiei unui tulburări de personalitate, cel mai potrivit ar fi ca cele două grupuri să fie alcătuite în pereche, din cadrul aceleiaşi familii, pentru că astfel se controlează din start o multitudine de variabile precum nivelul de educaţie al părinţilor, stilul parental, mediul cultural, tipul de şcoală urmată etc.). Dacă studiul urmărit nu permite utilizarea unor grupuri de control interne, alternativa ar fi aceea de a selecta un eşantion reprezentativ din populaţia generală care corespunde criteriilor respective. Acesta poate fi păstrat intact, aşa cum a fost el selectat prin tehnici de eşantionare probabilistică (cu avantajul de a păstra cel mai bine caracteristicile generale ale populaţiei de interes) sau poate fi modificat (se pot selecta subeşantioane cu un nivel ridicat de similaritate pe baza unor algoritmi cum este scorul compozit de propensitate descris în cadrul secţiunii 5.5) (cu avantajul de a avea un nivel sporit de similaritate cu grupul-ţintă). De

altfel, unul dintre aspectele obligatorii de prezentat în orice studiu pe cohorte este o descriere a caracteristicilor demografice şi referitoare la variabile potenţial confundate a loturilor comparate, astfel încât să se observe echivalenţa acestora la momentul iniţial.

***Analiza cazurilor lipsă (a participanţilor pierduţi*) (distorsiuni cauzate de pierderea participanţilor)**

Dacă procentul de pierdere a participanţilor din studiile longitudinale ia valori scăzute (de exemplu, până în 15%) nu sunt afectate fundamental rezultatele studiului, deşi pe măsură ce procentul de pierdere creşte, scade capacitatea de generalizare a rezultatelor la populaţia din care a fost extrasă cohorta. Ar fi însă de interes să se analizeze profilul persoanelor care au renunţat şi în ce măsură acest lucru nu ar distorsiona rezultatele (de exemplu, dacă majoritatea celor care au fost pierduţi din studiu aparţin grupului persoanelor dependente de alcool, se poate ca efectul alcoolului asupra comportamentului infracţional să fie subevaluat).

***Includerea şi măsurarea adecvată a variabilelor confundate* (distorsiuni cauzate de problema variabilei terţe)**

Includerea variabilelor potenţial confundate ar trebui făcută în baza unei analize atente a literaturii de specialitate, etapă premergătoare celei de iniţiere a studiului de colectare a datelor iniţiale. La acest nivel contează atât includerea tuturor variabilelor relevante, cât şi găsirea unor modalităţi adecvate de măsurare a acestora.

5.3.3. Aspecte statistice specifice în studiile de cohortă

Spre deosebire de celelalte tipuri de studii, bazate pe tehnici statistice de testare a ipotezelor (cu rezultate semnificative statistic la un $p < 0,05$), în cazul studiilor pe cohortă, îndeosebi a celor din domeniul clinic, se apelează la alte forme de analiză statistică. Prevalenţa, incidenţa, incidenţa cumulativă, riscul relativ, riscul atribuibil sunt concepte specifice întâlnite în studiile de acest gen, ele fiind prezentate succint în cele ce urmează, pentru a creşte familiarizarea publicului consumator de psihologie cu aceste concepte.

Înainte de a ilustra calculul acestora, se cuvine să facem câteva precizări de ordin terminologic. Prin prevalenţă se înţelege procentul de cazuri de interes identificate în populaţia testată (de exemplu, 9% din populaţia de adolescenţi şcolarizaţi consumă droguri uşoare ilegale). Prin incidenţă, se înţelege procentul de cazuri *noi* de interes identificate în populaţia testată raportat la o

anumită perioadă de timp (de exemplu, în fiecare an există 2,5% cazuri noi de adolescenţi care ajung să consume droguri), în calculul incidenţei nefiind luaţi în calcul cei care consumă de mai mult timp decât perioada aflată în urmărire. Din acest motiv, pentru că incidenţa se referă la numărul de cazuri noi apărute într-o perioadă de urmărire, în timp ce prevalenţa se referă la numărul total de cazuri observat în perioada urmărită, incidenţa ia valori mai mici decât prevalenţa. De asemenea, incidenţa poate fi estimată adecvat doar în cadrul unor studii longitudinale cu eşantion identic (studii pe cohortă sau panel), în timp ce prevalenţa[15] poate fi estimată inclusiv în cadrul unor studii transversale în care se asigură o reprezentativitate a eşantionului analizat la nivelul populaţiei de interes.

Pentru ilustrare, vom apela la situaţia ipotetică a unui studiu longitudinal cu două măsurători, realizate la 1 iunie 2012, respectiv 31 mai 2013 pe 1.000 de adolescenţi chestionaţi cu privire la consumul de marijuana. Cu ocazia primei testări, 60 dintre ei au declarat că au consumat ocazional asemenea droguri, iar cu ocazia celei de-a doua testări, alţi 30 au declarat că au început să consume droguri, în medie cu 6 luni înaintea celei de-a doua măsurări. De asemenea, 10 dintre cei 60 care au declarat iniţial că au consumat marijuana au specificat că au renunţat la această practică, iar alţi 10 fac parte dintre cei pierduţi la a doua evaluare. Cu ocazia celei de-a doua testări au fost disponibili spre a fi chestionaţi doar 920 dintre adolescenţii cuprinşi iniţial în studiu, 80 (8%) dintre ei fiind pierduţi pentru cea de-a doua examinare. Schema logică a acestui exemplu este prezentată în figura 5.4.

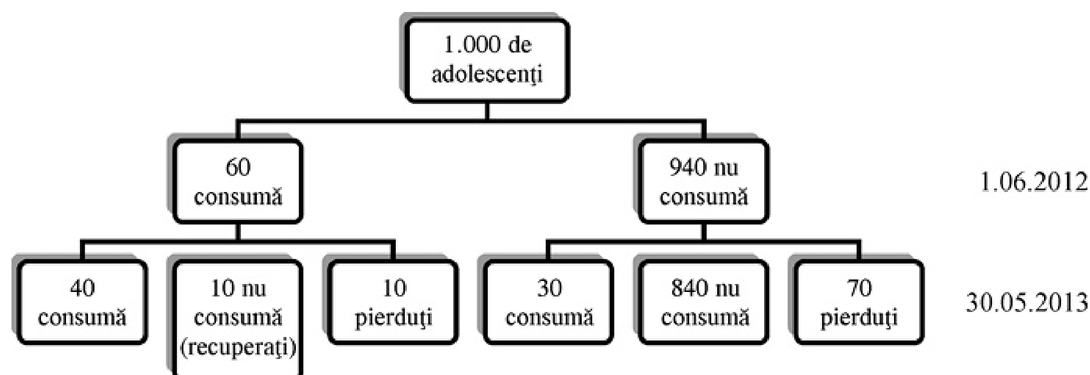


Figura 5.4. *Rezultatele prevalenţei şi incidenţei consumului de marijuana la adolescenţi*

Prevalenţa la 1.06.2012 – 6% (60 de cazuri la 1.000 de persoane)

Prevalenţa la 31.05.2013 – 7,6%, (40 + 30)/(40 + 10 + 30 + 840), adică 70/920

Prevalenţa anuală (între 01.06.2012 şi 31.05.2012) – 9%, (60 + 30)/1.000[16]

Incidenţa cumulativă (numărul de cazuri noi în intervalul de timp analizat) = 3,44%, 30/(940 – 70)

Rata incidenţei (densitatea incidenţei) = 3,50%/35 din 1.000 de persoane/anual, 30/([0,5 × 30] + [1 × 840]) sau 30/855[17]

Relaţia dinamică dintre incidenţă şi prevalenţă poate oferi perspective multiple de interpretare. De exemplu, dacă pe o cohortă urmărită în timp prevalenţa tinde să ia valori din ce în ce mai mari (de exemplu, şomajul să crească de la 6% la 9% din populaţie) în timp ce se observă o rată a incidenţei relativ constantă (numărul cazurilor de şomaj noi să fie relativ acelaşi), poate fi semnul că avem de-a face cu perioade tot mai lungi de şomaj, această stare de fapt fiind stimulată sau întreţinută fie de un context tot mai dificil de a găsi un alt loc de muncă, fie de o motivaţie mai scăzută a şomerilor de a căuta un alt loc de muncă (vezi partea dreaptă a figurii 5.5).

În schimb, luând acelaşi exemplu al şomajului, scăderea în timp a prevalenţei fenomenului, însoţită de o rată a incidenţei relativ constantă (scade rata şomajului, deşi procentul de şomeri noi este acelaşi), reflectă scurtarea perioadei de şomaj, această arătând fie măsuri economice benefice care să faciliteze reîncadrarea rapidă în muncă, fie un fenomen ascuns precum migrarea şomerilor către alte ţări (deşi incidenţa fenomenului este aceeaşi, prevalenţa măsurată la momente ulterioare nu identifică aceste cazuri pierdute prin emigrare) (vezi partea stângă a figurii 5.5).

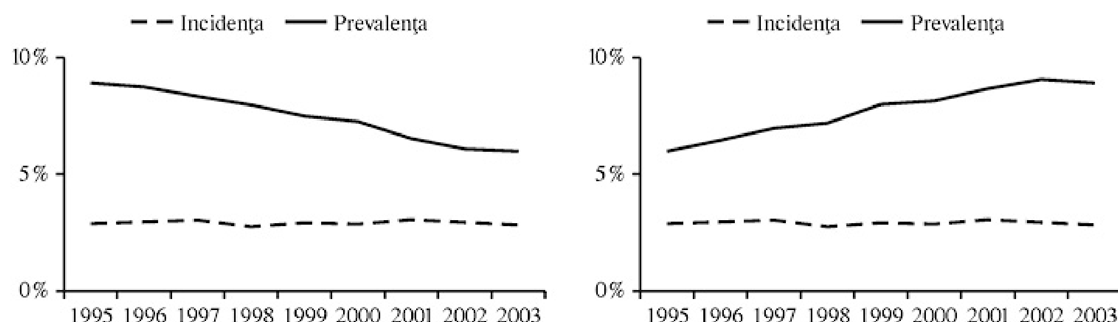


Figura 5.5. *Ilustrarea dinamicii dintre incidenţa şi prevalenţa unui fenomen*

O altă categorie de concepte relevante în zona studiilor efectuate pe cohorte vizează calcularea riscului relativ, a riscului atribuibil, respectiv a fracţiunii riscului (WHO, 2001). Acestea pot fi surprinse atunci când atât factorul de risc, cât şi efectul urmărit sunt operaţionalizaţi într-o manieră categorială (nominală dihotomică – de exemplu, depresie majoră prezentă sau absentă). Astfel, într-un studiu pe cohortă sunt urmărite un număr de 1.000 de persoane proaspăt căsătorite, dintre care 400 au fost investigate ca având tulburări de personalitate, iar celelalte 600 au o personalitate în registrul normalităţii. Eşantionul este urmărit tip de 10 ani, iar la finalul perioadei se verifică câţi păstrează aceeaşi relaţie de acum 10 ani (nu au divorţat). Rezultatele obţinute la finalul perioadei sunt surprinse în tabelul 5.7.

Tabelul 5.7. *Situaţia relaţiei dintre tulburările de personalitate şi divorţ*

		Statut civil după 10 ani		
		A divorţat	Nu a divorţat	Total
Tulburările de personalitate	Cu tulburări	120 (a)	280 (b)	400
	Fără tulburări	60 (c)	540 (d)	600
	Total	160	840	1.000

Riscul relativ[18] (RR) = (120/400)/(60/600) = 0,3/0,1 = 3

Riscul atribuibil (RA) = 120/400 – 60/600 = 0,3 – 0,1 = 0,2

Fracţiunea riscului (FR) = (RR – 1)/RR = 2/3 = 0,66

Rezultatele pot fi interpretate astfel: (a) există de trei ori mai multe şanse să se observe un divorţ dacă cel puţin unul dintre soţi suferă de tulburări de personalitate, decât dacă niciunul dintre aceştia nu are asemenea tulburări (riscul relativ); (b) 20% (0,2) din cazurile noi de divorţ apărute în grupul expus pot fi puse pe seama tulburărilor de personalitate (risc atribuibil); (c) 66% dintre divorţuri în cadrul celor diagnosticaţi cu tulburări de personalitate poate fi pus pe seama acestora (fracţiunea riscului).

5.3.4. Tehnici statistice specifice în studiile longitudinale

Modelul anterior de analiză statistică este întâlnit îndeosebi în zona clinică, fiind preluat treptat în domeniile psihologiei clinice şi a psihoterapiei. Însă tehnicile statistice întâlnite în studiile longitudinale pot fi extrem de diversificate în funcţie de o serie de aspecte precum: (1) caracteristica măsurătorilor prin care se urmăreşte schimbarea în timp; (2) rolul timpului – predictor sau criteriu (efect) urmărit.

De exemplu, cu referire la primul aspect menţionat, în studiile longitudinale care apelează la variabile numerice atât la nivel de predictori, cât şi la criterii (efecte) urmărite se utilizează analize statistice sofisticate, precum modelarea curbei de creştere a variabilei latente (*latent variable growth curve modeling*), modelarea multinivelară liniară ierarhică (*HLM – linear hierarchical modeling*) sau analiza tranziţiei latente ca formă a modelului clasei latente (*latent class model*). Analizele statistice depăşesc, prin complexitatea lor, posibilitatea de a fi discutate aici, însă caseta 5.4 prezintă pe scurt logica analizei statistice în funcţie de întrebarea de cercetare vizată, pornind de la câteva texte introductive precum Singer şi Willett (2003), Duncan, Duncan şi Strycker (2006).

Caseta 5.4. *Apelul la tehnici statistice în funcţie de întrebarea cercetării din studiul longitudinal*

Autorii unui studiu sunt interesaţi să evidenţieze cum evoluează comportamentul antisocial la pubertate şi adolescenţă. Pentru aceasta adoptă o strategie retrospectivă, selectând un eşantion aleatoriu de 500 de elevi din populaţia şcolară de clasa a XI-a, de la care obţin acceptul pentru a consulta arhivele cabinetelor de psihologie din şcoală, deoarece aceştia au fost evaluaţi anual, în mod obligatoriu, cu o scală care măsoară conduita antisocială. De asemenea, ei apelează la o serie de date sociometrice existente, colectate când cei studiaţi erau în clasa a III-a, pentru a identifica în ce măsură aceştia erau respinşi de către colegii lor, respectiv consideraţi a fi agresivi în baza unui test sociometric. În sfârşit, cu acordul elevilor şi al părinţilor acestora, autorii obţin date de la poliţie pentru a vedea în ce măsură adolescenţii studiaţi au comis vreo abatere soldată cu contravenţie sau vreo infracţiune, înregistrată în evidenţele poliţiei.

Cercetătorii şi-au pus următoarele patru întrebări:

(A) *Cum evoluează nivelul comportamentului antisocial pe măsură ce elevii cresc?* (întrebare cu caracter descriptiv – creştere sau descreştere liniară, stagnare etc.). Scopul acestei întrebări este de a descrie forma traiectoriei de creştere pentru fiecare individ în parte, cât şi identificarea patternului cel mai frecvent întâlnit în grupul studiat.

(B) *Poate fi prezisă evoluţia diferită a comportamentului antisocial* (unii devin mai agresivi, alţii îşi reduc manifestările antisociale etc.) *în funcţie de anumiţi predictori* (cum ar fi genul persoanei [diferenţe între băieţi şi fete] sau nivelul lor de integrare şcolară [copii percepuţi ca fiind agresivi şi respinşi de către colegi în clasa a III-a; copii agresivi, dar nerespinşi; copii acceptaţi şi neagresivi] etc.)*?* (întrebare descriptiv-analitică ce vizează relaţia dintre predictori şi patternul schimbărilor observate). Scopul acestei întrebări este de a surprinde măsura în care anumiţi predictori sunt asociaţi cu diferenţele interindividuale observate în traiectoria modificărilor observate.

(C) *Pe baza predictorilor utilizaţi se poate estima dacă şi când vor apărea probleme de comportament sancţionate de către poliţie?* (întrebare analitică cu rol predictiv).

(D) *Pe baza predictorilor utilizaţi şi evoluţiei nivelului comportamentului antisocial până în clasa a XI-a, se poate prezice cum va evolua acest comportament până când elevii vor împlini 25 de ani?* (întrebare analitică cu rol predictiv).

Primele două întrebări sunt abordate standard prin HLM. Acesta este un demers ierarhic ce conţine nivelul 1 de analiză (descrierea schimbărilor observate la nivel intraindividual), respectiv nivelul 2 de analiză (descrierea asocierii dintre predictori şi diferenţele observate în traiectoria schimbărilor indivizilor – abordare interindividuală). Pentru nivelul 1 de analiză este nevoie de un minimum de 3 valuri de măsurare (dar preferabil ar fi să existe cel puţin 5 valuri de măsurare) pentru a testa eficienţa diferitelor forme ale traiectoriei schimbării (liniare, neliniare, exponenţiale etc.). De asemenea, ar fi nevoie ca efectul să fie măsurat printr-un instrument cu proprietăţi psihometrice forte bune pentru toate categoriile de vârstă la care este aplicat sau care să demonstreze o bună invarianţă metrică. În privinţa selectării predictorilor, aceştia trebuie să respecte aceleaşi caracteristici prezentate anterior pentru studiile pe cohorte.

HLM este o tehnică robustă pentru a fi aplicată atunci când modificările sunt continue. Dacă schimbarea ia o formă stadială (creştere atât cantitativă, cât şi calitativă prin salturi – vezi exemplul stadiilor dezvoltării cognitive ale lui Piaget), atunci mai potrivit ar fi apelul la modelul claselor latente.

Cea de-a treia întrebare necesită o analiză a supravieţuirii (*survival analysis*, denumire ce-şi are originea în utilizarea ei în medicină, pentru a estima dacă şi pentru cât timp anumite proceduri medicale conduc la prelungirea vieţii, fiind cunoscută şi sub denumirea de *event-history analysis*). Tehnica este utilă în cadrul oricărui studiu longitudinal realizat pentru a estima dacă şi când va apărea evenimentul urmărit (în acest caz – prima abatere/infracţiune a adolescenţilor care este sancţionată de către poliţie).

În sfârşit, ultima întrebare ar putea fi adresată prin analiza seriilor de timp (*time-series analysis*), utilizată frecvent în domeniul economic pentru a face predicţii despre cum vor evolua anumiţi parametri economici pe baza datelor colectate în mod repetat de la alţi parametri care corelează cu variabila de interes.
Acest tip de analiză va încerca să prezică ce se va întâmpla în viitor, nu şi de ce se va întâmpla acel lucru, pornind de la premisa că evoluţia din viitor va urma aceleaşi tendinţe cu cele observate în trecut (evoluţie constantă, evoluţie ascendentă sau descendentă, fluctuaţii sezoniere, evoluţie ciclică), fiind potrivită îndeosebi pentru a face predicţii pe termen scurt. Una dintre cele mai utilizate modele de analiză a seriilor de timp este ARIMA (*Auto-Regressive Integrated Moving Average*).

5.4. Studiile corelaţionale

Probabil că cel mai întâlnit tip de studiu în domeniul psihologiei academice este cel bazat pe un design non-experimental ce urmăreşte obiective descriptiv-analitice din zona asocierii dintre variabile, fie ele studii corelaţionale (între două variabile măsurate numeric sau ordinal), fie studii ce vizează contingenţa (între două variabile măsurate nominal). Ele sunt atât de răspândite, încât l-au determinat pe Cronbach (1957) să afirme că existe două tipuri de psihologie, cea experimentală şi cea corelaţională. Prima este interesată de a studia şi explica varianţa intergrup (diferenţele dintre grupuri ca urmare a unei manipulări experimentale), în timp ce psihologia fundamentată pe corelaţii este interesată să surprindă diferenţele intragrupale (varietatea

indivizilor din cadrul aceluiaşi grup). Prima este preocupată să controleze prin design efectul potenţialelor variabile confundate, iar pentru aceasta este dispusă să analizeze fiinţa umană în laborator, până la unităţi care par lipsite de orice privire de ansamblu (impactul unui anumit stimul simplu asupra unei arii specifice corticale), în timp ce a doua analizează reacţiile indivizilor în mediul lor natural şi caută mai degrabă să surprindă variabilitatea în date, nu să o controleze. Iar atunci când este nevoie de un anumit control al variabilelor, acesta este exercitat prin tehnici statistice, nu prin design, ceea ce a dus la o dezvoltare semnificativă a tehnicilor statistice specifice studiilor corelaţionale.

Datorită numărului impresionant de studii corelaţionale, tehnicile statistice de bază pentru studiul asocierii datelor (coeficientul de corelaţie liniară simplă – Bravais-Pearson, respectiv tehnica hi pătrat pentru tabele de contingenţă) sunt predate studenţilor de la psihologie încă din primul an de studii. De aceea, este foarte probabil ca tuturor cititorilor acestui capitol să le fie familiară ideea că valorile lui r Pearson sunt cuprinse între –1 şi +1, unde o valoare 0 sau foarte apropiată acesteia indică absenţa unei corelaţii între variabilele analizate.

Ceea ce pare mai puţin clar, uneori, este partea de interpretare a rezultatelor. Una dintre frazele esenţiale de reţinut în domeniul cercetării psihologice este *corelaţia nu înseamnă cauzalitate*. Pentru a ilustra acest aspect, voi relua exemplul prezentat în Sava (2004/2011, prezentat în ediţia a II-a), pornind de la analiza lui Mitchell şi Jolley (2006) cu privire la patru posibile direcţii de interpretare a unei corelaţii negative semnificative statistic între stima de sine a unei persoane şi timpul alocat vizionării programelor TV.

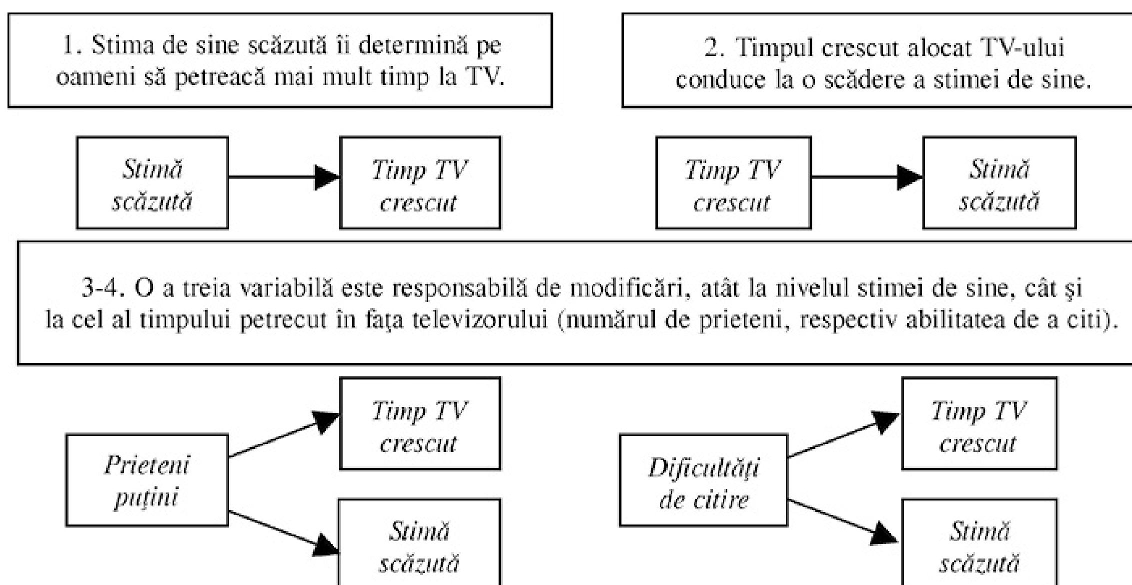


Figura 5.6. *Explicaţii alternative ale unei corelaţii negative între stima de sine şi timpul petrecut în faţa televizorului*

Aceste potenţiale explicaţii nu epuizează însă toate alternativele existente. De pildă, este posibil ca variabila terţă, responsabilă pentru corelaţia dintre timpul alocat emisiunilor TV şi stima de sine, să fie statutul socioeconomic al indivizilor. Persoanele cu un statut socioeconomic mai scăzut pot avea un nivel mai scăzut al stimei de sine şi pot petrece mai mult timp în faţa televizorului, această modalitate de petrecere a timpului liber fiind una dintre cele mai ieftine.

Aceste exemple reflectă cele două probleme majore ale oricărui studiu corelaţional: (a) problema direcţiei asocierii (care dintre variabilele asociate este o variabilă-cauză şi care este o variabilă-efect) (cazurile 1 şi 2 din figura 5.6); (b) problema celei de-a treia variabile (fie ea o variabilă mediatoare, moderatoare sau o variabilă care conduce la o relaţie falsă între două variabile – ultima fiind reflectată prin cazurile 3 şi 4 din figura 5.6). La aceste neajunsuri, trebuie avute în vedere şi riscurile la nivelul validităţii statistice (vezi tabelul 3.5 din capitolul 3).

Date fiind toate aceste probleme legate de interpretarea rezultatelor statistice obţinute ar fi oportun să ne întrebăm: „cu ce ne pot fi de folos?”,

respectiv „de ce sunt atât de răspândite?”.

Un prim argument în favoarea studiilor corelaţionale, şi a celor de asociere în general, priveşte tocmai problema cauzalităţii. Pe cât de adevărat este faptul că existenţa unei corelaţii nu semnifică neapărat o relaţie de tip cauză-efect, la fel de adevărat este faptul că asocierea este o condiţie a existenţei unei relaţii cauzale. În absenţa unei corelaţii între două variabile, putem fi siguri (dacă nu sunt probleme de validitate statistică sau dacă cei investigaţi nu se află la acel nivel al unei variabile moderatoare în care corelaţia dintre cele două este nulă) că nu există o relaţie cauzală între variabile. Una dintre condiţiile necesare, dar nu şi suficiente, ale demonstrării unei relaţii de tip cauză-efect este să existe o asociere între presupusă variabilă-cauză şi presupusa variabilă-efect. Prin urmare, studiile corelaţionale pot fi primul pas care trebuie realizat într-un demers complex de demonstrare a unei relaţii cauzale, deoarece presupun costuri financiare şi de timp mult mai mici decât studiile experimentale sau decât cele non-experimentale longitudinale.

În al doilea rând, în contexte aplicative pot exista situaţii în care cercetătorii să fie interesaţi doar de evidenţierea unei asocieri între două variabile, deoarece pe baza acestei relaţii se pot realiza predicţii în direcţia de interes. De exemplu, într-un context de selecţie de personal este important să se cunoască că există o asociere între conştiinciozitate şi performanţa în muncă, deoarece prin intermediul acestei relaţii putem estima care va fi nivelul viitor al performanţei cu ajutorul evaluării nivelului de conştiinciozitate. Această informaţie este valoroasă pentru orice specialist în resurse umane, chiar dacă nu este demonstrată o relaţie de tip cauză-efect între cele două variabile amintite.

De asemenea, studiile corelaţionale sunt potrivite pentru a răspunde la întrebări relativ timpurii cu privire la un anumit fenomen, atunci când încă nu se cunosc suficient de multe despre modul în care o anumită variabilă de interes (de exemplu, angajamentul organizaţional) corelează cu alte variabile specifice domeniului respectiv (de exemplu, intenţia de a părăsi organizaţia;

nivelul de epuizare profesională, performanţa în muncă, autonomia în muncă etc.).

În sfârşit, dată fiind complexitatea fenomenelor psihice, studiile de tip corelaţional sunt utile pentru a surprinde relaţii de tip multivariat (în care se urmăreşte asocierea simultană a mai mult de două variabile). Acest lucru l-a determinat pe acelaşi Cronbach (1957) să afirme că, probabil, cel mai valoros aspect adus de studiile corelaţionale este concepţia multivariată despre lume. În acest context este relevant să amintim de relaţiile de mediere, de relaţiile de moderare, de modelele complexe, cunoscute şi sub denumirea de modele structurale (de exemplu, analize de cale) sau, impropriu, de modele cauzale (vezi detalii în secţiunea 5.4.2)[19].

Aşadar studiile corelaţionale îşi dovedesc relevanţa într-o multitudine de situaţii. Wegener şi Fabregar (2000) enumeră trei dintre acestea: (a) dacă obiectivele cercetării, de tip descriptiv-analtic, vizează evidenţierea asocierii dintre variabile; (b) dacă nu este posibilă manipularea unei variabile cauză din punct de vedere etic sau pragmatic; (c) pentru a reduce costurile şi a creşte viteza de colectare a datelor, atunci când situaţia cere acest lucru. La acestea aş adăuga ideea subliniată de Cronbach (1957). Studiile corelaţionale sunt foarte utile atunci când se doreşte a se ţine seama de complexitatea relaţiilor psihologice, de multidimensionalitatea acestora (de exemplu, în studii care testează legătura dintre opt variabile prin intermediul unor modele structurale complexe sau relaţia dintre două seturi de variabile prin intermediul unei corelaţii canonice), indiferent dacă obiectivul vizat este unul explorator, unul descriptiv pur sau unul descriptiv-analitic (ce presupune testarea unor ipoteze de cercetare).

Cei mai mulţi specialişti, atunci când au în vedere studiile corelaţionale se gândesc la varianta de bază a acestui design non-experimental, asocierea dintre două variabile prin apelul la coeficientul de corelaţie Bravais-Pearson. În realitate există o multitudine de studii corelaţionale, diferenţiate în funcţie de aspectul principal urmărit (Wegener şi Fabrigar, 2000): (a) studii care apelează

la tehnici statistice ce vizează exclusiv aspecte necauzale; (b) studii care apelează la tehnici statistice ce îşi propun să vizeze anumite componente necesare unei relaţii cauzale (deşi nu pot răspunde în întregime la problematica cauzalităţii). Fiecărei categorii din aceste studii corelaţionale le vom rezerva câte o secţiune, pentru a descrie elementele fundamentale. Nu este nici obiectivul lucrării de faţă, nici spaţiul necesar pentru a intra în detalii, însă cititorii interesaţi pot găsi numeroase lucrări de statistică în spaţiul românesc, care să aprofundeze aceste teme (Sava, 2004/2011; Popa, 2008; Labăr, 2008; Culic, 2004).

5.4.1. Studii corelaţionale focalizate pe aspecte necauzale

O posibilă clasificare a diferitelor variante de studii care pun în prim-plan ideea corelaţiei existente între două sau mai multe variabile are în vedere obiectivul studiului, şi implicit, tehnica statistică utilizată. Tabelul 5.8 prezintă sintetic principalele tipuri de studii întâlnite.

Tabel 5.8. *Principale tipuri de studii bazate pe asocierea variabilelor*

Variaţii ale studiilor fundamentate pe corelaţii	*Descriere succintă*
Analiza factorială exploratorie – EFA (vezi detalii în Sava, 2004/2011)	Obiectivul acestui tip de studii este de a identifica şi descrie factorii (variabilele latente) care pot explica un anumit pattern corelaţional observat într-un set de variabile. Cea mai răspândită aplicaţie a acestui tip de analiză se întâlneşte în studii derulate în faza incipientă de dezvoltare a unor probe psihologice.
Analiza factorială confirmatorie – CFA (vezi detalii în Sava, 2004/2011)	Acest tip de studiu reprezintă o extensie a analizei factoriale exploratorii, fiind o analiză factorială care testează modele ce conţin anumite restricţii (de exemplu, ca un item să coreleze cu un anumit factor şi să fie independent de ceilalţi factori etc.). Analiza poate fi utilizată în fazele ulterioare ale dezvoltării unei probe sau ale unui model conceptual, când avem deja modele teoretice care să explice patternul corelaţional al datelor pe care dorim să le testăm. Cazuri

	particulare ale CFA sunt studiile bazate pe analiza matricei multitrăsătură-multimetodă (MTMM, *multitrait-multimethod model*), respectiv modele circumplexe.
Scalarea multidimensională (analiza de corespondenţă) (vezi detalii în Sava, 2004/2011 sau Culic, 2004)	Analiza se bazează pe reprezentarea grafică a similarităţii, respectiv a disimilarităţii dintre stimuli într-o hartă cognitivă (conceptuală). Cu cât două aspecte sunt mai puternic similare, cu atât ele sunt reprezentate grafic mai apropiat unul de celălalt pe o hartă. Dacă stimulii sunt măsuraţi prin scale numerice, se apelează la scalare multidimensională, iar dacă stimulii sunt măsuraţi într-o manieră nominală dihotomică, se apelează la analiza de corespondenţă.
Analiza de cluster (vezi detalii în Culic, 2004 şi o ilustrare în Sava şi Popa, 2011)	Analiza de cluster este utilă atunci când se doreşte clasificarea cazurilor observate în anumite categorii (tipologii) sau când se doreşte identificarea unor subcategorii (tipuri) psihologice dintr-o populaţie eterogenă (de exemplu, tehnica este foarte utilă în demersul de segmentare a consumatorilor).
Analiza de regresie în scop predictiv (vezi detalii în Sava, 2004/2011)	Analiza de regresie în scop predictiv este utilizată pentru a estima evoluţia unei variabile de interes, denumită criteriu, în funcţie de valorile unuia sau mai multor predictori (variabile care corelează cu criteriul). Pe baza acestei analize se poate prognoza, de exemplu, care va fi performanţa la bacalaureat a unui elev dacă se cunosc rezultatele şcolare ale acestuia din clasa a XI-a, nivelul său de inteligenţă generală şi nivelul său de anxietate de performanţă în situaţia unui examen.
Corelaţia canonică (vezi o ilustrare în Sava, 2009)	Corelaţia canonică reprezintă o extindere a situaţiei abordate prin analiza de regresie. Dacă prin utilizarea regresiei se poate observa evoluţia unui singur criteriu în funcţie de un set de variabile predictor, în corelaţia canonică poate fi analizată relaţia dintre un set de predictori şi un set de criterii (mai multe criterii).

Desigur, explicitarea specificului fiecărui studiu prezentat în tabelul 5.8 ar necesita probabil sute de pagini, de aceea ne-am rezumat la a face trimitere la unele referinţe bibliografice existente în spaţiul românesc, întâlnite preponderent în cărţi de statistică/analiză a datelor.

5.4.2. Studii corelaţionale focalizate pe surprinderea unor aspecte cauzale

Pe parcursul întregii cărţi am afirmat că existenţa unei corelaţii nu constituie o garanţie pentru existenţa unei relaţii cauzale între cele două variabile asociate. Prin urmare, am susţinut că nu se pot demonstra relaţii de tip cauză-efect prin studii corelaţionale. Însă acest lucru nu înseamnă că studiile corelaţionale nu pot fi utilizate pentru a surprinde anumite condiţii incluse printre criteriile considerării unui anumite relaţii ca fiind una de tip cauzal. Astfel, condiţia de bază pentru a surprinde o relaţie de tip cauză-efect este existenţa asocierii dintre variabila-cauză şi variabila-efect. Această condiţie necesară, dar nu şi suficientă, pentru a surprinde relaţii cauzale poate fi adresată prin studii corelaţionale, indiferent de natura scopului acestora.

Însă unele studii corelaţionale se focalizează nu numai pe condiţia asocierii dintre variabile, ci şi pe efortul de a surprinde o asemenea asociere după ce a fost eliminat efectul potenţialelor variabile confundate, aspect cunoscut sub denumirea de control statistic al variabilelor. Această abordare este frecvent întâlnită în studiile non-experimentale, deoarece în acest tip de studii nu se poate controla prin design impactul unor variabile (manipularea VI), prin urmare se compensează această limită prin efortul de a controla statistic al acestor variabile.

Mai mult decât accentul pus pe surprinderea unei asocieri, în condiţiile controlării la nivel statistic al efectului unor variabile potenţial confundate, un anumit tip particular de studii corelaţionale – corelaţiile încrucişate (*cross-lagged correlations*) sunt utilizate pentru a adresa întrebări legate de relaţia temporală dintre variabilele asociate, cunoscându-se faptul că modificările la nivelul variabilei-cauză ar trebui să preceadă temporal modificările observate la nivelul variabilei-efect. Toate aceste situaţii sunt reflectate succint în figura 5.7.

Fiecare dintre aceste tehnici de lucru – (a) corelaţiile încrucişate; (b) regresia ierarhică; (c) aşa-numitele „modele cauzale”, fie ele manifeste (analiza de cale) sau latente (modelarea prin ecuaţii structurale – SEM) şi (d)

apelul la variabile instrumentale – sunt prezentate succint în paginile următoare.

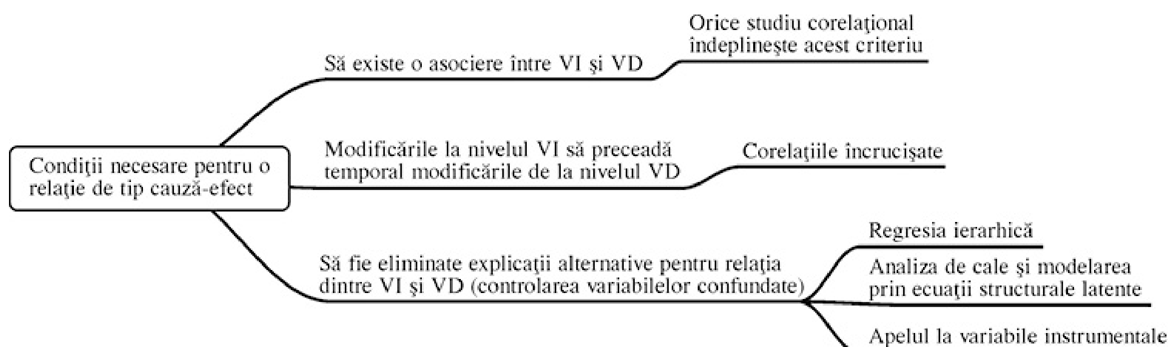


Figura 5.7. *O ilustrare a relaţiei dintre principalele condiţii necesare pentru a investiga o relaţie cauzală şi tipurile de strategii corelaţionale de analiză şi/sau design*[20]

5.4.2.1. Corelaţiile încrucişate

Corelaţiile încrucişate (*cross-lagged-panel correlation*) ar fi putut fi prezentate în rândul studiilor de tip longitudinal datorită designului pe care îl implică, fiind vorba de măsurători repetate. Am decis să le includem aici, deoarece implică studii corelaţionale care încearcă să ofere indicii asupra direcţiei unei asocieri observate, astfel încât să putem să apreciem care dintre cele două variabile corelate ar putea fi variabila-cauză şi care este variabila-efect. Pentru a înţelege logica acestei proceduri voi apela la un exemplu clasic oferit de Elmes, Kantowitz şi Roediger III (2012) pornind de la un studiu din anii ’70 de Eron şi colaboratorii.

Aceştia au studiat un lot de copii de clasa a II-a, privind atitudinile lor faţă de emisiuni (programe) cu conţinut violent la TV şi nivelul lor de agresivitate (măsurat indirect prin evaluare colegială – chestionarea colegilor cu privire la nivelul de agresivitate observat la cel evaluat). Corelaţia obţinută a fost una pozitivă şi semnificativă statistic (r = 0,21), ceea ce indică existenţa unei asocieri între preferinţa pentru emisiuni cu un conţinut violent şi nivelul lor de agresivitate. Din punctul de vedere al interpretării cauzale, nu putem însă să

stabilim dacă cei care sunt agresivi preferă să urmărească emisiuni cu mai multă violenţă sau dacă emisiunile violente conduc la o creştere/stimulare a agresivităţii. Pentru a rezolva această dilemă, acelaşi lot de participanţi a fost investigat 10 ani mai târziu, când elevii erau în clasa a XII-a, rezultatele fiind prezentate în figura 5.8. Aici sunt trecute toate corelaţiile posibile (de exemplu, între nivelul agresivităţii din clasa a II-a şi cel înregistrat în clasa a XII-a, cazul b, există o corelaţie pozitivă egală cu 0,38). Cele mai importante corelaţii de interpretat sunt cei doi coeficienţi de corelaţie asociaţi relaţiilor încrucişate (e şi f). Se observă că între nivelul agresivităţii din clasa a II-a şi preferinţa pentru conţinut violent la TV din clasa a XII-a (linia întreruptă) există o corelaţie de 0,01, ceea ce indică absenţa asocierii dintre cele două variabile. În schimb, între preferinţa pentru conţinut violent la TV din clasa a II-a şi nivelul agresivităţii din clasa a XII-a (linia îngroşată) există o corelaţie semnificativă pozitivă de 0,31. Interpretarea în tandem a celor două rezultate sugerează faptul că un consum de material media timpuriu conduce la creşterea nivelului de agresivitate. Interpretarea opusă, aceea că persoanele agresive par a conduce la urmărirea mai frecventă a programelor TV cu conţinut violent nu se susţine.

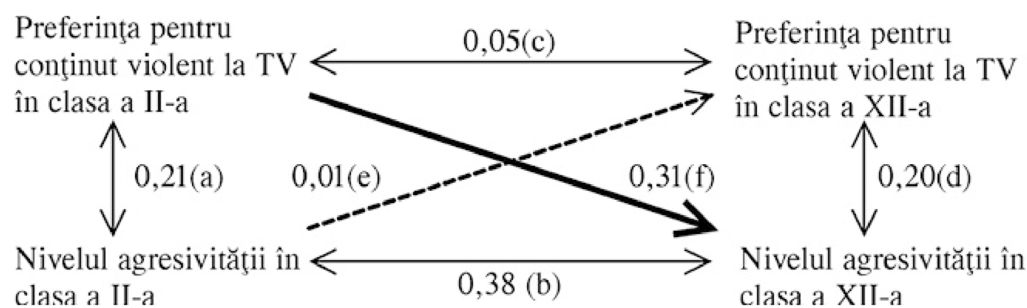


Figura 5.8. *O ilustrare a designului ce solicită analiza corelaţiilor încrucişate*

Desigur, nu întotdeauna rezultatele sunt la fel de simplu de interpretat. De exemplu, pot apărea contexte în care ambele corelaţii e şi f sunt semnificative statistic şi apropiate ca valoare, ori situaţii în care niciuna dintre ele nu este semnificativă statistic etc. Acest lucru îngreunează posibilitatea de a interpreta

rezultatele obţinute, mai cu seamă că pot exista şi cazuri în care: (1) relaţia la momentul T0 indică o asociere (corelaţia a); (2) interpretarea în tandem a rezultatelor la corelaţiile e şi f poate da indicii clare privind direcţia cauzalităţii; (3) dar corelaţia de la momentul T1 (corelaţia d) să fie una nesemnificativă statistic, fapt care ar anula întregul demers logic de identificare a unei relaţii cauzale.

Mai mult, chiar dacă cele două condiţii de bază ar fi îndeplinite: (a) corelaţii semnificative statistic pentru cazurile a şi d; (b) rezultat semnificativ statistic pentru e, dar nu şi pentru f sau viceversa, tot nu am putea fi siguri că am identificat corect direcţia relaţiei cauzale. Prin designul lor, corelaţiile încrucişate se adresează îndeosebi criteriul asociativ şi cel al temporalităţii pentru a realiza inferenţe cauzale, însă nu este la fel de bun pentru a elimina potenţiale explicaţii alternative determinate de variabile confundate (criteriul izolării asocierii). În concluzie, acest tip de studiu corelaţional, deşi oferă o perspectivă plauzibilă cu privire la inferenţele de tip cauzal, nu poate înlocui studiile experimentale, fiind mai puţin acurat în capacitatea sa de adresa o relaţie de tip cauză-efect.

5.4.2.2. Regresiile ierarhice

Studiile experimentale au marele merit de a controla impactul diferitelor variabile confundate asupra variabilei dependente prin egalizarea efectului lor ca urmare a alocării randomizate a participanţilor în condiţiile experimentale. Se realizează, aşadar, un control al variabilelor străine prin design. Prin apelul la tehnica regresiei ierarhice[21], efectul potenţialelor variabile confundate este controlat la nivel statistic. De exemplu, într-un studiu în care se doreşte urmărirea efectului cauzal al inteligenţei fluide asupra performanţei în muncă pe un lot de muncitori, se poate apela la o strategie în paşi de a verifica acest lucru. De exemplu, se poate compara care este procentul variaţiei scorurilor la performanţa în muncă care poate fi prezis de alţi factori (experienţa în muncă,

nivelul de cunoştinţe profesionale, motivaţia pentru muncă etc.), în două situaţii distincte, fără a include, respectiv după includerea inteligenţei în ecuaţia de regresie. O eventuală creştere semnificativă sesizată a puterii de explicare a variaţiei criteriului în ecuaţia de regresie în care a fost adăugată inteligenţa poate însemna că impactul inteligenţei asupra performanţei în muncă este unul semnificativ statistic.

Spre deosebire însă de procesul randomizării, care are rolul de a echilibra impactul potenţial al oricărei variabile pe care un cercetător l-a luat sau nu în calcul, în cazul controlului statistic, cum este cazul regresiilor ierarhice sau al corelaţiilor parţiale, pot fi controlate numai efectele variabilelor luate în calcul şi măsurate. Acest lucru nu garantează faptul că ar putea exista şi alte variabile confundate responsabile pentru creşterea sesizată în pasul al doilea, când a fost adăugată inteligenţa, aspect care ar conduce la concluzii eronate. În consecinţă, nici studiile bazate pe regresie ierarhică nu constituie instanţe la fel de solide precum experimentele, pentru a identifica relaţii cauzale. Deşi se acordă o grijă deosebită pentru două dintre aspectele esenţiale de demonstrat (există o relaţie dintre predictor şi criteriu, respectiv controlul unor variabile confundate măsurate), nu există certitudinea că au fost luate în seamă toate variabilele confundate relevante. În plus, nici direcţia cauzalităţii nu este foarte clară, în absenţa unui design de tip experimental sau a unuia non-experimental longitudinal.

Caseta 5.5. *Când controlăm statistic variabilele potenţial confundate?*

Pentru ca o variabilă să fie considerată confundată, ea trebuie să îndeplinească trei cerinţe: (1) să existe o asociere între factorul de risc şi aceasta; (2) să contribuie la apariţia efectului; (3) să nu facă parte din relaţia cauzală (distală) dintre factorul de risc şi efect (Bruce, Pope şi Stanistreet, 2008). De exemplu, o dietă inadecvată, bogată în grăsimi, poate fi o variabilă confundată în relaţia dintre fumat şi bolile cardiovasculare. Există o asociere pozitivă între fumat şi consumul de grăsimi (persoane puţin preocupate de aspectele de sănătate, în general) (1); ambele contribuie la creşterea apariţiei bolilor cardiovasculare (2); iar dieta bogată în grăsimi nu poate fi considerată ca elementul de legătură (variabilă mediatoare între fumat şi bolile coronariene) (3). În schimb, în relaţia dintre fumatul în timpul sarcinii şi riscul naşterii unui copil sub greutatea optimă, nivelul carbohemoglobinei din sânge nu poate fi

considerată o variabilă confundată, pentru că, deşi factorii (1) şi (2) sunt prezenţi (există o asociere între fumatul în timpul sarcinii şi nivelul carbohemoglobinei din sângele mamei, iar ambii factori contribuie la un risc ridicat de naştere a unor copii cu greutate mai mică decât cea normală), lipseşte cel de-al treilea element (3) (carbohemoglobina este o variabilă mediatoare – fumatul în timpul sarcinii conduce la creşterea nivelului de carbohemoglobină, care la rândul său este responsabil de apariţia unor naşteri de copii greutatea având sub valoarea optimă).

În general dorim să controlăm statistic efectul acestor variabilele confundate pentru a putea observa relaţia dintre variabilele de interes X şi Y nedistorsionată de aceşti factori străini. Aşa cum am menţionat în mod repetat pe parcursul acestei lucrări, sursa asocierii dintre două variabile poate fi de natură cauzală sau necauzală (determinată de anumite variabile confundate). Însă Lee (2012) prezintă o sursă suplimentară a existenţei unei asocieri dintre variabile, care este determinată artificial de situaţia controlării statistice a unei variabile confundate, când aceasta nu ar fi trebuit controlată. Astfel, poate apărea o corelaţie falsă între două variabile-cauză independente una de cealaltă, dacă se controlează statistic nivelul unei variabile-efect comune, ceea ce poate conduce la ideea unor rezultate fals pozitive, deşi au fost controlate anumite variabile relevante. Această variabilă care se constituie ca efect comun al altor variabile se numeşte în teoria grafurilor (Pearl, 2009) variabilă de ciocnire (*collider*). O ilustrare a acestei stări de fapt este oferită pornind de la relaţia inversă dintre nivelul stimei de sine şi timpul alocat

urmării emisiunilor TV ca mijloc de petrecere a timpului liber (vezi exemplele 3-4 din figura 5.6). Astfel, dacă variabilele confundate controlate sunt numărul de prieteni (reţeaua de suport social) şi dificultăţile de citire, vom ajunge probabil la concluzia că a dispărut asocierea între stima de sine şi timpul alocat emisiunilor TV (corelaţia iniţială fiind una falsă, determinată de faptul că ambele variabile aveau o cauză comună). Dacă însă vor fi cuprinse între variabilele confundate şi variabile precum singurătatea autopercepută (ca efect comun al unei stime de sine scăzute şi a unui timp crescut alocat urmăririi programelor TV), corelaţia iniţială va persista, lăsând impresia unei relaţii autentice între stima de sine şi consumul de programe TV.

Aşadar, în funcţie de relaţia dintre variabilele străine controlate statistic şi predictorii de interes, se poate ajunge la concluzii diferite privind rezultatul asocierii dintre predictori, după cum se poate vedea din situaţiile 2-4, acestea fiind cazurile posibile de extindere a situaţiei 1, prin exercitarea controlului statistic.

Situaţia 1. Consumul intens de programe TV prezice (relaţionează cu) stima de sine mai scăzută.

Situaţia 2. Consumul de programe TV nu mai prezice (relaţionează) cu stima de sine, dacă sunt controlate variabile antecedente considerate a fi cauze comune precum numărul (redus) de prieteni.

Situaţia 3. Consumul de programe TV prezice în continuare stima de sine, dacă sunt controlate variabile văzute drept consecinţe (efecte) comune ale stimei de sine şi a consumului de programe TV, cum ar fi nivelul de singurătate autoperceput.

Situaţia 4. Relaţia dintre consumul de programe TV şi stima de sine devine una incertă (dar poate fi în continuare una semnificativă statistic, chiar dacă avem de a face cu o corelaţie falsă) dacă sunt controlaţi concomitent atât factorii antecedenţi (numărul de prieteni), cât şi descendenţi (singurătatea autopercepută) în funcţie de intensitatea legăturii dintre aceste variabile potenţial confundate şi variabilele de interes.
Un alt exemplu oferit de Asendorpf (2012) cu referire la relaţia inversă dintre statutul de imigrant (ca variabilă *dummy*, grec emigrant în altă ţară din Europa – 1; grec care trăieşte în Grecia – 0) şi nivelul autoeficacităţii generale, dacă se controlează nivelul autoperceput de cunoaştere a limbii oficiale. Însă această variabilă poate fi suspectată ca fiind o variabilă de ciocnire deoarece atât ţara în care trăiesc, cât şi nivelul de autoeficacitate generală pot reprezenta factori cauzali plauzibili pentru nivelul autoperceput de cunoaştere a limbii oficiale din ţara respectivă[22].
Pentru a defini corect modelul predictiv testat (Lee, 2012) recomandă apelul la reprezentare grafică şi, pe cât posibil, utilizarea modelelor de ecuaţii structurale (descrise mai jos) în locul modelelor bazate pe ecuaţiilor de regresie.

5.4.2.3. Analiza de cale (cu variabile manifeste) şi ecuaţiile structurale (cu variabile latente)

Dacă prin apelul la tehnica regresiei pot fi abordate doar situaţii cauzale directe (de exemplu, faptul că X este un factor cauzal plauzibil pentru Y, în contextul în care am controlat efectul lui Z1 şi Z2 asupra lui Y), analiza de cale şi ecuaţiile structurale latente permit discutarea unor modele „cauzale” complexe. De exemplu, ele permit diferenţierea dintre variabile care joacă rolul unei cauze distale (X îl determină pe Y, prin intermediul lui M) şi cele care au rolul unei cauze proximale (M îl determină pe Y) (vezi relaţia de mediere, cazul B, din figura 5.9). De asemenea, ele pot urmări măsura în care X îl determină pe Y în mod diferit, în funcţie de nivelului lui Z (Z jucând rolul unei variabile moderatoare, cazul C din figura 5.9). Acest tip de raţionament poate fi extins spre a verifica modele „cauzale” complexe, prin apelul la analiza de cale sau ecuaţii structurale latente (vezi cazul D din figura 5.9). Diferenţa dintre analiza de cale (*path analysis*) şi ecuaţiile structurale latente (*structural equation modeling* – SEM sau *covariance structure modeling*) este una de natură psihometrică. În analiza de cale se lucrează cu un sigur indicator

al unei variabile manifeste (de exemplu, rezultatul direct de la testul Raven pentru a reprezenta conceptul de inteligenţă). În cazul ecuaţiilor structurale latente se lucrează cu variabile latente constituite în urma măsurării a cel puţin trei variabile manifeste pentru fiecare construct evaluat instrumentului (de exemplu, se trec rezultatele de la testul Raven, testul de desen al omuleţului şi de la testul WISC ca indicatori manifeşti ai variabilei latente – inteligenţa). Aceste măsurători multiple ale aceluiaşi construct conduc la eliminarea erorilor de măsurare ale unui construct, determinate de aspecte particulare ale unei probe şi obţinerea unor valori teoretice cu privire la intensitatea asocierii dintre constructele pe care le reprezintă.

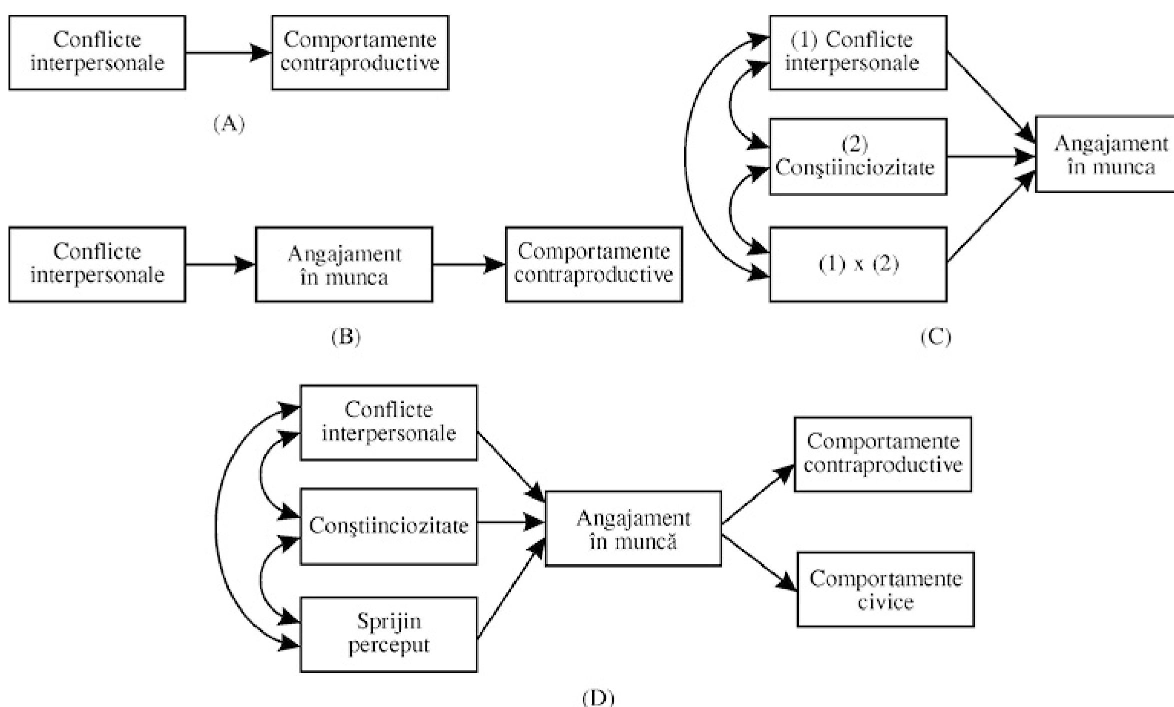


Figura 5.9. *Tipuri de relaţii cauzale testate prin analiză de cale: (A) relaţii directe, (B) relaţii indirecte (efecte mediate), (C) relaţii condiţionate (efecte moderate), (D) relaţii complexe*

Relaţiile complexe (cazul D) sunt însă ţinta principală de interes în studii care apelează la analiză de cale sau la ecuaţii structurale latente, deoarece toate celelalte cazuri pot fi rezolvate şi prin intermediul regresiei, prin anumiţi paşi complecşi de lucru[23].

De aceea, pentru a simplifica lucrurile, analizele de cale sau modelarea prin ecuaţii structurale sunt mai potrivite decât analizele de regresie pentru a testa modele structurale complexe (de exemplu, cele care cuprind mai mult decât o singură variabilă mediatoare şi/sau mai mult decât un efect analizat). Bunăoară, exemplul D din cadrul figurii 5.9 este un model testat de Sulea şi colaboratorii săi (2012) pentru a urmări în ce măsură angajamentul în muncă mediază (poate explica) nivelul la care o serie de variabile personale şi organizaţionale (resurse precum conştiinciozitatea sau sprijinul perceput din partea şefilor, dar şi factori de presiune precum nivelul conflictual între angajaţi) afectează implicarea angajaţilor în comportamente dezirabile (civic-participative) sau indezirabile (contraproductive). Testarea unui asemenea model poate viza relaţii specifice (de exemplu, legătura dintre conştiinciozitate şi angajamentul în muncă), însă ea se centrează pe testarea întregului model, acest lucru reprezentând, de altfel, principalul punct de interes atunci când se apelează la astfel de modele.

Un dezavantaj al modelelor complexe este ceea ce aş putea numi problema balanţei dintre modelul structural cel mai bun şi modelul *parcimon* (simplificat). De exemplu, testarea modelului D din figura 5.9 poate conduce la rezultate statistice considerate a fi bune, însă dacă modelul ar fi modificat puţin, prin stabilirea unei legături directe între, să zicem, conflictele interpersonale şi comportamentele contraproductive, indicatorii statistici vor deveni şi mai buni. Desigur, pe măsură ce creăm mai multe legături între variabile, modelul se complică, iar indicatorii statistici devin tot mai buni. Scopul nu este însă de a trasa legături între toate variabilele (toate combinaţiile posibile), ci de a găsi cel mai simplu model (cu cât mai puţine legături stipulate), care să aibă indicatori statistici suficient de buni.

O altă limită importantă derivă din relaţia autentică dintre variabile. Simplu fapt că am trasat grafic o relaţie cauzală între două variabile (de exemplu, între conflicte percepute între colegi şi angajamentul în muncă) nu asigură veridicitatea relaţiei. Este posibil ca angajamentul în muncă să fie mai

degrabă o variabilă-cauză care să conducă la mai puţine situaţii conflictuale între colegi. De aceea, ar fi util să fie comparate mai multe modele concurente, pentru a afla care dintre ele prezintă mai bine realitatea observată (patternul de corelaţii dintre variabile). În sfârşit, ca în orice relaţie de tip cauzal, rămâne problema faptului ca modificarea VI să preceadă temporal schimbările de la nivelul VD, aspect ce poate fi adresat adecvat prin apelul la un design longitudinal. Unele modele avansate de tip SEM pot avea în vedere inclusiv designuri de tip longitudinal.

5.4.2.4. Variabilele instrumentale

Apelul la variabile instrumentale pentru a susţine inferenţe de tip cauzal este destul de răspândit în domeniul ştiinţelor economice (Bollen, 2012; Crosby *et al.*, 2010), dar aproape complet absent în sfera psihologiei. Prin variabilă instrumentală se înţelege acea variabilă care este asociată cu efectul urmărit doar prin intermediul potenţialei variabile independente. De exemplu, dacă problema studiată este efectul fumatului (ca posibilă variabilă-cauză) asupra apariţiei cancerului (variabila-efect), avem mai mare încredere în legătura cauzală dintre acestea dacă se observă o asociere între cancer şi nivelul accizelor stabilite pentru consumul de tutun (aceasta fiind o variabilă instrumentală deoarece afectează consumul de tutun şi este neconectată logic cu efectul studiat). Un alt exemplu este relaţia dintre nivelul de şcolarizare (cauză) şi venitul obţinut după absolvire (efect), dacă luăm în calcul distanţa faţă de unitatea de învăţământ drept variabilă instrumentală (distanţele mai scurte favorizând perioadele mai lungi de şcolarizare, fără a fi conectate logic de veniturile obţinute, decât prin intermediul numărului de ani de şcoală).

Iată logica de lucru atunci când apelăm la variabile instrumentale. Fie Z – variabila instrumentală, X – variabila cauză, Y – variabila efect şi U – alte variabile incluse în ecuaţie (covariabile). O variabilă poate fi considerată instrumentală dacă: (a) Z este independent de U, (b) Z este asociat cu X destul

de puternic[24] şi (c) Z devine nesemnificativ statistic în relaţie cu Y dacă se controlează nivelul lui X şi a lui U[25]. Astfel, variabila instrumentală este asociată efectului observat doar prin intermediul potenţialei variabile cauză (variabila independentă ca o variabilă mediatoare). Ea nu corelează cu alte variabile potenţial confundate pentru relaţia dintre X şi Y.

Efectul variabilei potenţial cauză (predictorului) asupra variabilei-efect (criteriului) se testează printr-un tip particular de regresie (*two-stage least square* sau *2SLS*), implementat de majoritatea programelor statistice generale, inclusiv SPSS, după verificarea în prealabil a îndeplinirii condiţiilor de a considera o variabilă ca fiind instrumentală. În primul pas se estimează nivelul variabilei-cauză pe baza valorilor variabilei instrumentale, utilizată ca predictor, iar valorile prezise sunt salvate ca o variabilă nouă. În cel de-al doilea pas, variabila-efect este estimată pe baza valorilor prezise ale variabilei-cauză obţinute în pasul 1, iar valoarea coeficientului beta este interpretat într-o manieră clasică.

Pentru a ilustra mai bine rolul variabilelor instrumentale, vom apela la exemplul unei cercetări care urmărea să identifice impactul pe care anii de şcoală îl au asupra nivelului veniturilor (relaţie pe care mulţi români obişnuiţi o percep stereotipic a fi nesemnificativă statistic). O posibilă variabilă confundată între numărul de ani de şcoală şi venitul realizat este inteligenţa, deoarece oamenii inteligenţi tind să facă mai mulţi ani de şcoală, dar şi să se descurce mai bine în viaţă şi să câştige mai mulţi bani. Ingeniozitatea apelului la o variabilă instrumentală este aceea de a găsi o variabilă care să fie asociată cu lungimea şcolarizării, dar care să nu fie relevantă nici pentru nivelul de inteligenţă, nici pentru câştigurile financiare în sine, în acest ultim caz fiind posibilă o asociere doar dacă ea este mediată de lungimea şcolarizării. Variabila instrumentală pe care Angrist şi Krueger (1991) au utilizat-o cu referire la o populaţie din SUA a fost perioada din an în care s-a născut copilul. Cei născuţi în prima parte a anului pot rămâne încă un an acasă, astfel încât să înceapă şcoala mai târziu decât cei născuţi în a doua parte a anului.

Pentru toată lumea, însă, există o prevedere legislativă conform căreia şcoala este obligatorie în SUA până la împlinirea vârstei de 17 ani. Acest lucru însemna în cazul celor care părăseau şcoala la 17 ani că cei născuţi în prima perioadă a anului aveau în medie mai puţine luni de şcoală decât cei născuţi în a doua perioadă a anului, diferenţa dintre cele două loturi fiind aproape de un an de zile, pentru că primii începeau şcoala la o vârstă mai mare (media fiind de 7,5 ani faţă de 6,8 ani). O asemenea variabilă instrumentală ar putea fi ideală dacă se dovedeşte că ea corelează destul de puternic (sau măcar moderat) cu lungimea şcolarizării, dar care din punct de vedere logic nu este legată nici de inteligenţă, nici de veniturile realizate ulterior (decât, în acest ultim caz, printr-o relaţie mediată de lungimea perioadei de şcolarizare). Ea îl ajută pe cercetător să aprecieze măsura în care lungimea şcolarizării are un impact real (potenţială cauză) asupra veniturilor, dacă la finalul aplicării regresiei 2SLS se obţine o valoare beta semnificativă între lungimea perioadei de şcolarizare şi venituri.

5.5. Studiile comparative *ex post facto*

Un alt tip de studii non-experimentale destul de răspândit sunt studiile comparative cu grup de control, cunoscute în limba engleză şi sub denumirea de *case-control studies*[26]. Modelul standard presupune compararea a două grupe de interes care nu sunt formate ca urmare a manipulării VI de către cercetător, ci sunt construite, cel mai frecvent, în baza unor criterii naturale. Exemple de astfel de studii ar fi: (a) compararea anxietăţii postnatale a mamelor care au născut prin metoda naturală *vs* cezariană; (b) urmărirea eventualelor diferenţe în nivelul de angajament organizaţional între angajaţii din mediul privat şi cei bugetari; (c) identificarea performanţelor şcolare la teste standardizate internaţionale între elevii din sistemul tradiţional de

învăţământ şi cei proveniţi din şcoli alternative cum ar fi pedagogia Waldorf etc. Acest gen de studii comparative se numesc comparaţii *ex post facto*.

Spre deosebire însă de studiile de licenţă bazate pe clasicele comparaţii dintre două grupe, fiecare alcătuite din câte 30 de subiecţi, în studiile comparative bine documentate, se pune un accent deosebit pe demonstrarea echivalenţei grupelor comparate cu referire la alte variabile relevante pentru efectul urmărit decât cea utilizată pentru alcătuirea grupelor. De exemplu, dacă se compară atitudinea femeilor din mediul rural cu a celora din mediul urban faţă de utilizarea prezervativelor, ar trebui să ne asigurăm că cele două loturi comparate nu diferă în privinţa altor variabile relevante pentru subiectul abordat cum ar fi nivelul de educaţie, nivelul de religiozitate, statutul lor marital etc.

Principalul algoritm prin care se asigură această echivalenţă a scorurilor este apelul la un scor compozit de propensitate (*propensity score*). Acest scor reprezintă, de fapt, probabilitatea unui individ de a face parte din grupul de interes (Katz, 2010), de exemplu, femei din mediul rural, date fiind anumite caracteristici măsurate cum ar fi nivelul de studii, vârsta, nivelul de religiozitate etc. În cazul studiilor randomizate cu două grupe, această probabilitate este 0,50, ceea ce înseamnă şanse egale ca individul să facă parte din cele două grupe în urma alocării aleatorii. În cazul studiilor nerandomizate, probabilitatea nu este aceeaşi. De exemplu, există o probabilitate mai mare ca o femeie să facă parte din grupul celor din mediul rural dacă a absolvit studii elementare şi are un nivel mai ridicat de religiozitate, decât dacă are studii superioare şi un nivel mai scăzut de religiozitate. Ideea din spatele utilizării scorului compozit de propensitate este aceea că, odată calculat, el permite un control statistic al diferenţelor dintre grupe fie prin potrivire (*matching*) sau prin stratificare, fie prin utilizarea acestui scor ca o covariabilă, pentru a controla eventualele diferenţe dintre grupe. Acest lucru este important, pentru că se poate obţine un control real al variabilelor confundate, asemenea situaţiei randomizării, dacă indivizii din cele două grupe comparate au scoruri de

propensitate similare (de exemplu, unei femei din mediu rural cu scor de propensitate egal cu 0,75, îi corespunde o femeie din mediul urban, cu scorul de propensitate de 0,75 sau foarte apropiat acestuia).

Modalitatea prin care se calculează aceste scoruri de propensitate este una extrem de simplă, prin apelul la regresie logistică binară, unde grupa de care aparţine subiectul este variabila-criteriu (de exemplu, femei din mediu rural = 1, femei din mediu urban = 0), iar alte variabile relevante care pot distinge între cele două loturi, precum nivelul de educaţie, nivelul de religiozitate, statutul marital, vârsta la care a survenit prima căsătorie etc., sunt utilizate ca predictori pentru a obţine ecuaţia de regresie necesară pentru a calcula probabilitatea ca ea să aparţină grupului de femei din mediu rural pentru fiecare persoană. Detalii despre modul de calcul al regresiei logistice în literatura autohntonă sunt disponibile în Sava (2004/2011).

După obţinerea scorurilor de propensitate pentru fiecare individ din grupurile de interes, următorul pas este selectarea indivizilor cuprinşi în comparaţia finală. De exemplu, dacă ne dorim să comparăm femeile din mediul rural cu cele din mediul urban cu privire la atitudinea lor faţă de utilizarea prezervativelor ca mijloc de contracepţie, este posibil să strângem rezultate de la 200 de femei din mediul rural şi 200 de femei din mediu urban.

O posibilă soluţie constă în potrivirea celor 200 de femei din mediul rural cu 200 de femei din mediul urban (constituirea de perechi pe baza scorului de propensitate, astfel încât, unei femei care are scorul egal cu 0,36 într-un grup, să i se găsească o pereche în celălalt grup cu aceeaşi valoare sau una similară). Această metodă asigură cel mai ridicat nivel de echivalenţă între cele două grupe, însă prezintă un risc crescut de a pierde unii subiecţi (de exemplu, sunt excluse din analiză cazurile care nu şi-au găsit o pereche, pierderea fiind mai mare mai ales dacă grupele comparate sunt formate dintr-un număr inegal de subiecţi, o parte dintre aceştia rămânând fără pereche din start).

O altă posibilă soluţie constă în stratificarea indivizilor din cele două loturi. De exemplu, în cazul dat am putea gândi trei straturi de persoane:

indivizi cu scoruri de propensitate între 0,00 şi 0,29 (femei cu probabilitate redusă de a fi din mediul rural), indivizi cu scoruri medii (între 0,30 şi 0,70) şi indivizi cu scoruri ridicate, cuprinse între 0,71 şi 1,00 (femei cu probabilitate ridicată de a fi din mediul rural). Apoi cele două grupe sunt formate astfel încât să existe în fiecare un număr egal (sau procentual egal) de femei aparţinând fiecărui strat (de exemplu, 55 de femei cu probabilitate redusă de a proveni din mediul rural, în grupul celor din mediul rural şi 55 de femei cu probabilitate redusă de a proveni din mediul rural, incluse în grupul celor din mediul urban). Stratificarea prezintă un risc redus de pierdere a unor participanţi din studiu, dar are o putere mai scăzută de a asigura echivalenţa celor două grupe, comparativ cu potrivirea. Acest lucru este cu atât mai evident cu cât sunt constituite mai puţine straturi, fapt care ar permite o eterogenitate mai mare a probabilităţilor din cadrul unui strat (de exemplu, valori cuprinse între 0,30 şi 0,70).

În sfârşit, o alternativă posibilă ar fi să lăsăm grupurile aşa cum sunt ele constituite în cadrul studiului non-experimental, dar să utilizăm aceste scoruri de propensitate drept covariabile, pentru a controla statistic impactul diferenţelor iniţiale dintre grupuri, prin apelul la tehnici statistice precum ANCOVA.

Pentru a observa relevanţa utilizării scorului de propensitate în echivalenţa grupelor comparate, se poate observa în tabelul 5.9 consecinţa utilizării metodei potrivirii în perechi în baza scorurilor la propensitate, pentru exemplul dat.

Se observă că, în urma potrivirii cazurilor din cele două grupe comparate pe baza scorurilor de propensitate, deşi s-a pierdut un număr însemnat de indivizi din analiză, există un câştig important: similaritatea cazurilor privind alte variabile cu impact asupra efectului urmărit. În aceste condiţii, eventualele diferenţe observate cu privire la atitudinea faţă de prezervative între femeile din mediul urban şi cele din mediul rural nu mai pot fi puse pe seama altor

factori cum ar fi vârsta, nivelul de educaţie etc., spre deosebire de situaţia existentă înainte de potrivire.

Tabelul 5.9. *Caracteristici demografice ale participanţilor la studiu privind diferenţele dintre femeilor din mediul urban şi cele din mediul rural cu privire la atitudinea faţă de prezervative ca mijloace de contracepţie*

Aspecte analizate	Înainte de potrivire			După potrivirea cazurilor		
	Urban N=200	Rural N=200	p*	Urban N=115	Rural N=115	p*
Vârsta participanţilor (media)	36,0	42,3	< 0,01	40,0	40,4	0,54
Nivel de educaţie (ani de şcoală) (media)	12,5	10,1	< 0,05	10,8	10,5	0,23
Religiozitate (media)	6,2	7,8	< 0,05	7,1	7,2	0,67
Număr parteneri sexuali în ultimii 5 ani (media)	1,8	1,5	0,18	1,6	1,6	0,99
Consum zilnic de alcool (procente)	25%	29%	0,51	27%	26%	0,78
Consum zilnic de tutun (procente)	43%	25%	< 0,01	30%	27%	0,37
Cursuri de educaţie sexuală urmate (procente)	30%	15%	< 0,01	19%	16%	0,24

* $p < 0,05$ indică diferenţe semnificative între cele două loturi comparate

Deşi am insistat atât de mult pe problematica utilizării scorurilor de propensitate pentru a reduce distorsiunea de selecţie dintre grupele comparate, trebuie subliniat faptul că acest procedeu nu se poate substitui ideii de alocare randomizată pentru un motiv foarte simplu. Prin alocarea randomizată sunt formate două grupe echilibrate[27] din perspectiva oricărei variabile avute în vedere sau trecută cu vederea, dintre cele care ar fi putut afecta validitatea internă a rezultatelor obţinute. În schimb, prin apelul la potrivire în baza scorurilor de propensitate pot fi controlate statistic doar variabilele incluse în ecuaţia de regresie logistică, cu alte cuvinte, doar predictorii avuţi în vedere. Este foarte posibil însă să existe variabile potenţial confundate care să scape analizei sau care să nu poată fi măsurate din motive obiective. În exemplul oferit, în urma potrivirii cazurilor, cele două grupe par echivalente, astfel încât eventualele diferenţe observate pot reflecta, într-adevăr, diferenţe la nivel

atitudinal cauzate de specificul cultural al mediului de rezidenţă (rural *vs* urban). Însă este, de asemenea, posibil ca diferenţele constatate să apară din cauza unei alte variabile confundate care nu fusese luată în calcul. O asemenea variabilă posibilă este aceea de a cunoaşte persoane concrete infectate cu HIV sau bolnave de SIDA printre vecini (în mediul rural, unde fiecare cunoaşte mai bine pe ceilalţi, 20% au afirmat acest lucru, în timp ce în mediul urban doar 5% au declarat că ştiu personal pe cineva infectat cu HIV/SIDA). Din acest punct de vedere, eventualele diferenţe semnificative sau nesemnificative statistic între cei din mediul urban şi cei din mediul rural nu mai pot fi înţelese ca fiind un rezultat al diferenţelor culturale, atâta timp cât cele două grupe diferă şi în ceea ce priveşte expunerea la consecinţele negative ale absenţei utilizării prezervativelor în timpul actelor sexuale.

Atunci când din anumite motive nu se poate apela la scoruri de propensitate pentru a facilita echivalenţa grupelor comparate, se poate apela la o soluţie compensatorie, aceea de a include în grupul de control persoane care împărtăşesc în mod natural cât mai multe caracteristici comune cu cei din grupul-ţintă vizat. De la celebrele studii realizate pe gemeni, exemplele pot continua în a cuprinde în grupul de control persoane care au acelaşi mediu cultural şi familial precum cei din grupul-ţintă vizat. De exemplu, dacă se urmăreşte surprinderea unor diferenţe în prevalenţa personalităţii de tip A la hipertensivi *vs* persoane care nu suferă de hipertensiune arterială, grupul de control poate fi format din rude apropiate ale bolnavilor hipertensivi, cu condiţia de a avea un nivel similar de vârstă şi de a locui în aceeaşi gospodărie. Pentru a controla sexul persoanei, putem balansa astfel încât să avem un număr egal de femei şi bărbaţi în fiecare dintre cele două loturi ş.a.m.d.

Problemele studiilor de acest gen nu se rezumă doar la grija pentru echivalenţa grupelor comparate. Şi în acest caz, ca şi în cel al studiilor experimentale, ar fi ideal să existe o definiţie a condiţiei urmărite (grupul-ţintă). De exemplu, într-un studiu în care se compară copiii cu ADHD (deficit

de atenţie şi hiperactivitate) cu cei fără ADHD în privinţa performanţei şcolare, ar trebui foarte clar definit ce înseamnă un copil cu/fără ADHD. Acest lucru ar permite selectarea cazurilor în grupul respectiv în baza unor criterii coerente pentru toată lumea (de exemplu, ar fi considerat inacceptabil ca în grupul celor cu ADHD să intre copii diagnosticaţi pe baza KID-SCID-ului alături de copii nominalizaţi de către diriginţi ca având probleme cu atenţia sau de către psihologi, în baza performanţei scăzute la un test de atenţie concentrată, fiind nevoie de un criteriu unitar).

În sfârşit, acest tip de studii, destul de răspândit în domeniul medicinei, poate fi utilizat ca un prim pas atunci când studiile longitudinale pe cohortă nu sunt posibile, pentru a identifica anumiţi factori predispozanţi pentru o anumită tulburare. Designul este util mai ales dacă tulburarea respectivă este una rară sau dacă ea necesită un timp îndelungat de latenţă (incubare), deoarece aceste situaţii pot fi surprinse mai greu prin studii longitudinale pe cohorte, derulate în perioade mai scurte de timp (de exemplu, timp de 3 ani). Aşadar, studiile comparative de tip caz-control nu se rezumă doar la situaţia comparării existenţei unei diferenţe între mediile celor două grupe comparate, ci poate fi folosit şi pentru compararea existenţei unei diferenţe la nivel de frecvenţe (vezi cazul ilustrat în tabelul 5.10).

Într-un asemenea context, se poate calcula raportul şanselor sau raportul hazardului (*odd ratio* sau *hazard ratio*), care reprezintă, de altfel, principalul indicator al mărimii efectelor (vezi capitolul 7) pentru compararea de frecvenţe.

Tabelul 5.10. *Ilustrarea relaţiei dintre tipul de şcoală primară urmată (de stat* vs *privată) şi performanţa la olimpiadele şcolare din gimnaziu (premiat* vs *nepremiat)*

	Mediul şcolar la care este expus		
	Şcoli private	Şcoli de stat	Total
Performanţa la olimpiadă			
Premiat	a(239)	b(40)	a + b = 279
Fără premii	c(3.023)	d(1.287)	c + d = 4.310

	a + c = 3.262	b + d = 1.327	n = a + b + c + d = 4.589

Formula de calcul este una simplă, raportul şanselor fiind egal cu a × d/b × c (adică 239 × 1.287/3.023 × 40 = 2,54) pentru cazul ilustrat în tabelul 5.10, ceea ce înseamnă că există de 2,5 ori mai multe şanse să avem un elev premiat la olimpiadă de la şcoli private decât de la şcoli de stat). Raportul şanselor egal cu 1 înseamnă şanse egale de a atinge rezultatul vizat (premiile) sau de a fi diagnosticat cu tulburarea urmărită (ADHD) în ambele condiţii de expunere (şcoli de stat *vs* şcoli private; familii monoparentale *vs* familii biparentale). Valorile supraunitare ale raportului şanselor indică şanse mai mari să apară elementul urmărit (premiul, ADHD diagnosticat) în condiţiile de expunere speciale (şcoli private, familii monoparentale) decât în grupul de control (şcoli de stat, respectiv familii biparentale). Valorile subunitare ale raportului şanselor conduc la o interpretare opusă, şansele de apariţie a rezultatelor (premii, respectiv cazuri de ADHD) fiind mai mici în condiţiile speciale de expunere (şcoli private, respectiv familii monoparentale) decât în grupurile de control. Cititorii interesaţi pot găsi şi alte detalii despre interpretarea raportului şanselor în Sava (2004/2011), în capitolul destinat prezentării regresiei logistice.

5.6. Studiile de diagnostic

În domeniul psihologiei, acest tip de studii este puţin dezvoltat, spre deosebire de domeniul ştiinţelor medicale unde a cunoscut o înflorire deosebită (Akobeng, 2006; Petrie şi Sabin, 2009). Am decis să prezint această categorie specifică de studii non-experimentale datorită valorii lor euristice ridicate pentru dezvoltarea domeniului psihologiei. De altfel, designul studiilor de diagnostic (cu rolul de a identifica clar un anumit aspect – prezenţa unui caz de tulburare depresivă majoră) poate fi preluat şi pentru a face predicţii

(prognoze) cu privire la evoluţia lucrurilor (vor exista sau nu tentative de suicid pentru cazurile analizate diagnosticate cu depresie). În domeniul organizaţional, pe baza unor studii de diagnostic similare (analiză/selecţie) se poate urmări apoi dacă candidaţii angajaţi vor confirma ulterior statutul lor de persoane care îşi desfăşoară eficient activitatea lor profesională (cu scop predictiv/prognostic).

În mod tradiţional, acest tip de studii sunt concentrate sub forma unor tabele dihotomice de tip 2 × 2 (vezi tabelul 5.11), pentru că în cele mai multe situaţii diagnosticul sau prognosticul formulat ia o formă categorială (tulburare prezentă *vs* absentă). În astfel de analize sunt introduse o serie de concepte specifice precum sensibilitatea, specificitatea, falşi pozitivi etc. (vezi tabelul 5.11).

Tabelul 5.11. *Ilustrarea calculării principalilor parametri într-un test de diagnostic/prognostic*

	Tulburare prezentă Diagnostic de ADHD prezent	Tulburare absentă Diagnostic de ADHD absent	
Test screening pozitiv SDQ ≥ 7	(a) 44 de cazuri identificate corect	(b) 6 cazuri de fals pozitiv	Total screening pozitiv (a + b) = 50
Test screening negativ SDQ < 7	(c) 10 cazuri de fals negativ	(d) 440 de cazuri identificate corect	Total screening negativ (c + d) = 450
	Total cazuri cu ADHD (a + c) = 54	Total cazuri fără ADHD (b + d) = 446	Total testări (a + b + c + d) = 500

Sensibilitatea: proporţia de copii cu tulburare ADHD, care obţin un rezultat pozitiv la screening – a/(a + c), adică 0,815.
Specificitatea: proporţia de copii fără ADHD, care obţin un rezultat negativ la screening – d/(d + b), adică 0,986.
Valoarea predictivă pozitivă: proporţia celor cu screening pozitiv care au tulburarea ADHD – a/(a + b), adică 0,880.
Valoarea predictivă negativă: proporţia celor cu screening negativ care nu suferă de ADHD – d/(d + c), adică 0,977.

Exemplul ales porneşte de la adaptarea unui studiu efectuat de Ullebø şi colaboratorii săi (2011) cu privire la utilitatea folosirii chestionarului SDQ (Goodman, 1997). Acest instrument este disponibil în regim gratuit şi în limba română (detalii pe www.sdqinfo.com). Chestionarul a fost conceput ca un instrument de screening pentru a identifica o serie de resurse psihologice interne, dar mai ales descoperirea dificultăţilor, printre care se regăsesc şi probleme legate de atenţie şi hiperactivitate. Problema ridicată în studiul efectuat de cercetătorii norvegieni în 2011 este în ce măsură acest chestionar de screening este eficient în a prezice cazurile care îndeplinesc criteriile de diagnostic psihiatric de ADHD la copiii din şcoala primară. Dacă instrumentul se va dovedi a fi eficient, se vor reduce costurile serviciilor de evaluare (prin aprofundarea evaluării doar în cazurile semnalate prin instrumentul de screening, mai cu seamă că datele colectate la nivel naţional indică o prevalenţă a cazurilor de ADHD la vârsta şcolară mică de 5% – 5 copii din 100 au ADHD). Rezultatele obţinute în urma acestui studiu sunt sintetizate în tabelul 5.11.

Pornind de la rezultatele prezentate în tabelul 5.11 se observă că 815 din 1.000 de copii cu ADHD obţin un test pozitiv prin SDQ, ca instrument de screening (sensibilitatea). Se au în vedere doar cazurile care sunt diagnosticate cu ADHD, de aceea sensibilitatea unui test de screening ne spune doar o jumătate de poveste – cât de bine poate instrumentul de screening identifica pe cei care prezintă o tulburare, dar nu ne spune nimic despre cei care nu prezintă ADHD.

Specificitatea (*specificity*) este conceptul care ne spune cealaltă jumătate din poveste cu privire la calitatea unui instrument de screening. Din tabelul 5.11 se observă că 986 din 1.000 de copiii care nu au ADHD obţin un test negativ la SDQ. Acest rezultat este deosebit de bun pentru că ne spune că testul de screening respectiv reuşeşte să aibă un număr foarte redus de cazuri de fals negativ dintre cei fără ADHD şi testaţi cu SDQ. Specificitatea nu ne spune însă nimic despre cât de bună este proba pentru a-i identifica corect pe

cei care au ADHD, ci doar că este o probă foarte bună de a-i identifica corect pe cei care nu au ADHD.

Pornind de la aceste precizări, se cuvine să semnalăm că un test cu sensibilitate ridicată este foarte bun pentru a-i identifica pe cei fără ADHD dacă testul screening este negativ. De ce se întâmplă acest lucru? Să presupunem că proba SDQ ar fi avut o sensibilitate de 0,99 (99%) şi o specificitate de 0,70 (70%). Dacă o persoană ar fi fost testată pozitiv la screening nu înseamnă automat că ea are ADHD, deoarece o sensibilitate de 99%, neînsoţită de o specificitate la fel de ridicată, înseamnă un număr destul de mare de falşi pozitivi (cu test screening pozitiv, dar fără ADHD). În schimb, o persoană cu un test de screening negativ este foarte puţin probabil să aibă ADHD, deoarece sensibilitatea ridicată a probei arată că 99% dintre cei cu ADHD obţin un rezultat pozitiv la screening, ceea ce înseamnă o probabilitate foarte mică pentru cineva cu test de screening negativ să prezinte ADHD.

În schimb, specificitatea ridicată este foarte potrivită pentru a identifica persoane cu ADHD dacă rezultatul la screening este unul pozitiv. Să presupunem că la aceeaşi probă SDQ ar fi avut o sensibilitate de 0,70 şi o specificitate de 0,99. Dacă o persoană ar fi fost testată negativ la screening, nu înseamnă automat că ea nu are ADHD, deoarece o specificitate de 99%, neînsoţită de o sensibilitate la fel de ridicată, înseamnă un număr destul de mare de falşi negativi (cu test screening negativ, dar cu ADHD). În schimb, în acest context, o persoană cu screening pozitiv este foarte probabil să aibă ADHD, deoarece o specificitate ridicată a probei arată că 99% dintre cei fără ADHD obţin un test screening negativ, ceea ce conduce la o probabilitate foarte mică ca cineva cu un test screening pozitiv să nu prezinte ADHD (probabilitate foarte mare să aibă ADHD).

Analiza datele prezentate în tabelul 5.11 (sensibilitate de 0,815 şi specificitate de 0,986) ne indică faptul că SDQ ca instrument de screening este util îndeosebi în situaţiile în care rezultatul la screening este unul pozitiv, fiind

foarte mare probabilitatea ca indivizii cu un test pozitiv să aibă ADHD (test cu specificitate foarte mare).

Un aspect teoretic foarte important legat de *sensibilitate* şi *specificitate* este acela că acestea fac *referire la calităţile testului* de a-i identifica corect pe cei care prezintă sau nu o anumită tulburare, însă nu pot fi utilizate direct pentru a aprecia probabilitatea unei persoane testate de a prezenta o anumită tulburare. Acestea pot fi estimate doar dacă se apelează la alte concepte relevante care iau în calcul deopotrivă valorile predictive pozitive, respectiv negative, precum şi prevalenţa tulburării în populaţia studiată (cât de răspândită este problema?).

Valoarea predictivă pozitivă (VPP) se referă la probabilitatea celor cu test de screening pozitiv să aibă într-adevăr tulburarea urmărită (ADHD) (dacă ne raportăm la exemplul dat, 88% dintre cei cu test screening pozitiv). Valoarea predictivă negativă (VPN) se referă la probabilitatea celor cu test screening negativ să nu aibă ADHD (dacă ne raportăm la acelaşi exemplu, 97,7% dintre cei cu test screening negativ). Desigur, din valoarea VPN se poate afla procentul celor cu test negativ care sunt diagnosticaţi, totuşi, cu ADHD (1 – VPN, în acest caz 2,3% dintre cei cu un test screening negativ au ADHD).

Aceste *valori VPP şi VPN se referă tot la test*, nu la probabilitatea unui anumit individ testat de a avea sau nu tulburarea respectivă (de exemplu, ADHD). Pentru a calcula probabilitatea unui individ cu un anumit rezultat la testul de screening (pozitiv sau negativ) de a fi diagnosticat cu ADHD se ţine cont, atât de aceste valori, cât şi de prevalenţa (gradul de răspândire) tulburării respective. În cazul unei tulburări cu prevalenţă ridicată, indivizii au o probabilitate mai mare să obţină un VPP ridicat, iar în cazul unei tulburări cu prevalenţă scăzută, să obţină un VPN ridicat.

În acest sens, în scopul calculării probabilităţii unui individ de a avea tulburarea analizată (de exemplu, ADHD), este nevoie mai întâi să introducem conceptul de raport al probabilităţii (LR – *likelihood ratio*). Astfel, vorbim despre un LR+ care se referă la raportul dintre probabilitatea cazurilor cu un

rezultat pozitiv la testul de screening care au tulburarea şi probabilitatea cazurilor cu un rezultat pozitiv la testul de screening care nu au tulburarea. De asemenea, vorbim despre un LR– care se referă la raportul dintre probabilitatea cazurilor cu un rezultat negativ la testul de screening, dar care au tulburarea şi probabilitatea cazurilor cu un rezultat negativ la testul de screening şi care nu au tulburarea[28].

Transpunând datele ilustrate în tabelul 5.11 rezultă un LR+ de 58,21, respectiv un LR– de 0,18. Interpretarea rezultatului lui LR+ indică faptul că o persoană cu ADHD are de aproape 60 de ori mai multe şanse să obţină un rezultat pozitiv la testul de screening decât o persoană fără ADHD. Interpretarea lui LR– indică faptul că o persoană cu ADHD are de peste 5 ori (1/0,18) mai puţine şanse de a obţine un rezultat negativ dacă are ADHD comparativ cu o persoană care nu are ADHD. În general, sunt recomandate valorile LR peste 10 sau sub 0,1 (Akobeng, 2007), ceea ce înseamnă că proba SDQ se dovedeşte a fi un instrument de screening eficient mai ales pentru a diferenţia între cei cu ADHD şi cei fără ADHD, în cazul unui rezultat pozitiv la screening.

Aceste valori ale lui LR+, respectiv LR– pot fi ajustate (ponderate) în funcţie de prevalenţa tulburării vizate, deoarece una este să ne referim la o tulburare rară în populaţia analizată (de exemplu, o prevalenţa de 2% din populaţie), iar altceva ar fi să ne referim la cazul unei tulburări frecvent întâlnite în populaţia analizată (de exemplu, 30%)[29,30]. Aceste valori ajustate nu exprimă probabilitatea existenţei unei tulburări, ci şansa ca aceasta să apară (şansa fiind egală cu raportul dintre probabilitatea de apariţie a tulburării şi probabilitatea ca ea să nu apară).

Pentru a nu apela la asemenea calcule complicate, se poate consulta vizual nomograma lui Flagan (*apud* Akobeng, 2007). În figura 5.10 sunt prezentate trei cazuri distincte, pentru o prevalenţă a unei tulburări în populaţie de 2%, de 10%, respectiv de 30%, cu precizarea că prin această metodă se face referire la valorile LR neponderate (de exemplu, un LR+ = 58,21 pentru exemplul dat).

Prin trasarea unei linii drepte din zona stângă spre dreapta se poate estima probabilitatea ca persoana testată să aibă tulburarea respectivă. Partea stângă a nomogramei indică prevalenţa unui fenomen (probabilitatea de apariţie a tulburării în absenţa vreunui indiciu cum este testul de screening, adică o probabilitate pretest). La mijlocul monogramei este specificată valoarea LR avută în vedere (în acest exemplu este vorba de LR+). Prelungind linia dreaptă până în zona dreaptă, se ajunge la a estima probabilitatea ca persoana testată (cu rezultat pozitiv la screening în cazul LR+) să aibă într-adevăr tulburarea respectivă. Se observă faptul că la aceleaşi valori ale LR+, probabilitatea ca o persoană să aibă ADHD este: (a) de peste 95% (aproximativ 97%) dacă prevalenţa ADHD în populaţia din care face parte individul testat este de 30%; (b) puţin sub 90% (aproximativ 87%) dacă prevalenţa ADHD este de 10%; (c) şi de aproximativ 55%, dacă prevalenţa ADHD este de 2%. O altă interpretare posibilă ar fi: (a) în cazul unui individ cu rezultat pozitiv la testul de screening, probabilitatea ca acesta să aibă ADHD creşte: (a) de la 30% la 97%; (b) de la 10% la 87%; (c) de la 2% la 55%.

Ca o observaţie, dacă testul de screening este eficient, probabilitatea unei tulburări faţă de media populaţiei va creşte în cazul unui rezultat pozitiv la screening, respectiv va scădea în cazul unui rezultat negativ la screening (vezi linia întreruptă pentru LR– de 0,18, se observă o scădere a probabilităţii de a avea ADHD de la 10% (prevalenţa fiind 10%), la o probabilitate sub 2%.

La finalul acestei secţiuni dedicate testelor de diagnostic/prognoză, ar fi util de precizat câteva lucruri despre conceptul de scor-limită (*cut off score*).

Prin scorurile-limită se are în vedere valoarea care diferenţiază în cadrul unui test de screening între un rezultat pozitiv şi unul negativ. În exemplul oferit, un scor mai mare sau egal cu 7 la itemii specifici ADHD din SDQ erau consideraţi ca fiind un rezultat pozitiv (risc ridicat de a fi prezentă tulburarea), în timp ce un scor mai mic decât 7 era considerat un rezultat negativ (cel mai probabil tulburarea de ADHD fiind absentă). Acestea nu se stabilesc arbitrar, ci de obicei sunt selectate astfel încât să maximizeze fie atât sensibilitatea şi

specificitatea unui test atunci când calitatea testului permite acest lucru, fie numai una dintre acestea atunci când calitatea testului nu permite maximizarea ambelor şi/sau când este mai importantă doar una dintre ele[31]. De exemplu, în cazul unei probleme cu o prevalenţă rară sau foarte rară (de exemplu, 1%), specificitatea poate fi considerată mai importantă deoarece este important să-i identifici corect pe cei puţini care au tulburarea respectivă în baza unui test pozitiv. Dacă este o problemă foarte gravă, cu un impact ridicat, s-ar putea pune mai mult accent pe sensibilitate, chiar dacă există un risc mai mare de apariţie a unor cazuri de fals pozitiv. De exemplu, în cazul unei boli grave precum cancerul este mai important ca testul de screening să poată să-i elimine corect pe cei care nu au această maladie, chiar dacă vom avea un număr mai mare de alarme false în urma screeningului (persoane cu test pozitiv, care ulterior se vor dovedi a nu avea cancer).

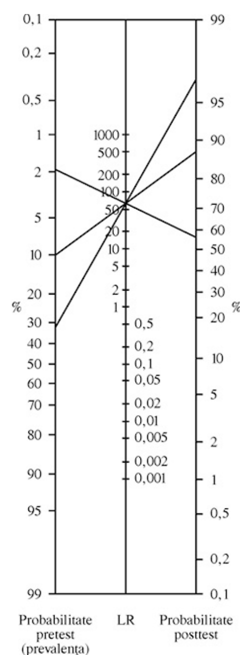


Figura 5.10. *Nomograma Flagan utilizată pentru a estima probabilitatea unei persoane de a avea problema vizată în funcţie de prevalenţa problemei şi de calitatea testului de screening*

Modalitatea standard de stabilire a acestor scoruri limită este prin intermediul analizei curbei ROC (disponibilă inclusiv în SPSS). O introducere

excelentă, relativ netehnică, a acestei analize este prezentată de Pintea şi Moldovan (2009).

Acest tip de analize se pretează foarte bine în orice domeniu al psihologiei, nu doar în sfera clinică. De exemplu, în baza unor probe psihologice scurte se poate face screeningul candidaţilor care au abilităţi cognitive normale sau deficitare în demersul de evaluare psihologică pentru a acorda un aviz de apt sau inapt psihologic, reducând considerabil timpul de lucru. În Sava (2004/2011) este analizat rolul unor asemenea instrumente de screening în domeniul testării psihoaptitudinale periodice din sfera psihologiei muncii. De exemplu, proba de atenţie concentrată Flanker AC (Sava şi Pătruţ, 2010), deşi are o durată de doar 3 minute, reuşeşte foarte bine să discrimineze pe cei care vor lua cu siguranţă avizul cu decizia apt psihologic de cei care primesc decizia de inapt sau pe aceea de apt cu anumite recomandări (rezerve) din motive aptitudinale. Printr-un asemenea demers s-a stabilit că un scor limită de 15 răspunsuri corecte sau mai puţin de 15 la Flanker AC este considerat un test pozitiv pentru prezenţa inaptitudinii cognitive, la o prevalenţă de 6% din cazuri şi un LR+ de 77,5, ceea ce înseamnă, raportându-ne la nomograma Flagan, că pentru un individ care obţine un asemenea scor creşte probabilitatea de a fi declarat la finalul evaluării inapt sau apt aptitudinal „condiţionat” de la 6% la 85%.

Aceeaşi logică poate fi aplicată şi în situaţii de selecţie, cu precizarea că rolul testului de screening este luat de evaluarea psihologică din cadrul selecţiei (în baza unui scor compozit), iar aspectul urmărit poate fi, de exemplu, fluctuaţia de personal (dacă după 1 an de zile de la angajare mai lucrează sau nu în compania respectivă). Un asemenea demers ne-ar ajuta să precizăm care este probabilitatea de a părăsi compania într-un interval mai mic de un an de zile, dacă prevalenţa anuală a acestui fenomen este de 15% pentru noii veniţi în firmă.

Concluzii

Demersurile non-experimentale acoperă o paletă largă de tipuri de studii care împărtăşesc toate aceeaşi caracteristică, aceea de a fi realizate fără ca cercetătorul să manipuleze variabilele de interes. În absenţa acestei manipulări este nepotrivit a se vorbi despre variabile independente şi cele dependente, fiind preferabile exprimări precum predictor şi criteriu, factori antecedenţi şi consecinţe, sau pur şi simplu variabile cuprinse în studiu.

Cele mai multe dintre studiile non-experimentale sunt potrivite pentru a adresa obiective exploratorii, descriptive şi descriptiv-analitice. Ele sunt mai puţin potrivite pentru a adresa întrebări de tip cauzal. Anumite aspecte ale cauzalităţii pot fi abordate îndeosebi prin studii de tip longitudinal (problema variabilei-cauză care precedă temporal efectele obţinute) sau prin studii corelaţionale complexe orientate spre controlul statistic al impactului unor variabile potenţial confundate (problema eliminării explicaţiilor alternative pentru efectul observat). Acest tip de studii sunt indicate fie în fazele timpurii ale abordării unei relaţii cauzale (pentru a fi utilizate ca argumente în vederea pornirii unui studiu experimental), fie în situaţii în care din anumite motive etice sau pragmatice, studiile experimentale nu sunt posibile.

Datorită unui accent mai redus pus pe controlul variabilelor prin design, studiile non-experimentale pot compensa printr-un control exercitat la nivel statistic. De aceea studiile non-experimentale apelează la un număr extrem de variat de tehnici statistice avansate. De altfel, o caracteristică compensatorie întâlnită este aceea că studiile non-experimentale apelează la analize multivariate, încercând să surprindă complexitatea patternurilor relaţionale complexe din realitate.

Diversitatea este un alt cuvânt-cheie care le caracterizează. În capitolul de faţă am selectat doar o parte dintre studiile de tip non-experimental, atât după criteriul răspândirii lor (vezi studiile corelaţionale), cât şi după potenţialul lor euristic care să conducă la dezvoltarea unor domenii, cum este cel al evaluării

psihologice (vezi studiile de diagnostic). Ele pot pune accent pe individ sau situaţii particulare (studiile de caz) sau pot viza aspectele generale, globale (sondajele). Ele pot pune accent pe metoda ştiinţifică utilizată pentru colectarea datelor (studii de caz – interviuri, sondaje – chestionare) sau pe metoda utilizată pentru analiza datelor (metode ce urmăresc asocierea – studii corelaţionale sau metode de comparare – studiile comparative *ex post facto*). De asemenea, ele pot urmări evoluţia unor variabile în timp (vezi studiile longitudinale pe cohorte) sau pot fi derulate transversal (majoritatea celorlalte studii non-experimentale). Aceste studii nu epuizează întreaga gamă de studii non-experimentale posibile. Dintre variantele de studii non-experimentale neincluse în acest capitol aş aminti aici studiile cu o metodologie calitativă (de exemplu, studiile fenomenologice) sau studiile metodologice, cum ar fi cele de validare şi/sau de adaptare culturală a unui test psihologic.

Lecturi suplimentare

Rotariu, T., Iluţ, P. (2006). *Ancheta sociologică şi sondajul de opinie*. *Teorie şi practică.* Iaşi: Editura Polirom.

Cartea este foarte utilă pentru cei interesaţi să aprofundeze problematica sondajului. Cartea conţine atât sfaturi practice despre cum trebuie construit un chestionar, cum trebuie selectat un eşantion, cum se pot evita anumite erori care pot distorsiona rezultatele, dar şi fundamente teoretice importante în sondarea opiniei publice.

Singer J.D., Willett, J.B. (2003). *Applied Longitudinal Data Analysis. Modeling change and event occurrence.* New York: Oxford University Press. Capitolele 1, 2, 3, 9 şi 10.

Cei care doresc să aprofundeze cunoştinţele cu privire la designul şi mai ales la analiza datelor provenite din studii longitudinale de diverse tipuri vor găsi răspunsuri la cele mai multe dintre întrebări. Nivelul tehnic al discuţiei este unul mediu, fiind una dintre puţinele lucrări care tratează acest subiect care îşi propune să explice cititorilor nefamiliari cu studiile longitudinale, dar care cunosc tehnici statistice de bază cum ar fi corelaţia sau regresia liniară.

Wegener, D.T., Fabrigar, L.R. (2000). Analysis and design for nonexperimental data: addressing causal and noncausal hypotheses. În H.T. Reis şi C.M. Judd (eds.), *Handbook of research methods in social and personality psychology* (pp. 412-450). New York: Cambridge University Press.

Capitolul recomandat trece în revistă o serie de studii corelaţionale complexe, precum analiza factorială exploratorie, analiza factorială confirmatorie, analiza de cluster, scalarea multidimensională, ecuaţiile

structurale latente, modelarea ierarhică multinivelară. Rolul capitolului este de a familiariza cititorii cu specificul acestor tehnici complexe, dar şi de a reliefa aspectele importante de care trebuie ţinut cont atunci când se face designul unui studiu care apelează la astfel de tehnice şi atunci când se doreşte să se interpreteze adecvat rezultatele obţinute cu ajutorul acestor tehnici.

Sava, F.A. (2011). Analiza datelor în cercetarea psihologică. Cluj-Napoca: Editura ASCR. Capitolele 6 şi 7 – „Regresia multiliniară" şi „Regresia logistică".

Capitolele respective introduc o serie de concepte relevante precum ideea de mediere, moderare, raportul şanselor, sensibilitate, specificitate, fals pozitiv, fals negativ, regresie ierarhică pentru controlul variabilelor potenţial confundate.

Bibliografie

Ahmed, S. (2009). *Statistical methods for sample surveys*. Găsit la 10.10.2012, disponibil la: http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/StatMethodsForSampleSurveys/coursePage/index/

Akobeng, A.K. (2006). Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Pædiatrica*, 96, pp. 487-491.

Akobeng, A.K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Pædiatrica*, 96, pp. 644-647.

Angrist, J., Krueger, A. (1991). Does Compulsory School Attendance Affect Schooling and Earnings?. *Quarterly Journal of Economics*, 106, pp. 979-1014.

Asendorpf, J.B. (2012). Bias due to controlling a collider: a potentially important issue for personality research. *European Journal of Personality*, 26, pp. 391-392.

Barlow, D.H., Nock, M.K., Hersen, M. (2008). *Single case experimental designs: Strategies for studying behavior change* (ediţia a III-a). Boston: Allyn & Bacon.

Baron, R.M., Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, pp. 1173-1182.

Belli, G. (2009). Non-experimental quantitative research. În S.D. Lapan M.T. Quartaroli (eds.), *Research Essentials: An Introduction to Design and Practices (Research Methods for Social Sciences)* (pp. 59-77). San Francisco, CA: Jossey-Bass.

Bollen, K.A. (2012). Instrumental variables in sociology and social sciences. *Annual Review of Sociology*, 38, pp. 37-72.

Bruce, N., Pope, D., Stanistreet, D. (2008). *Index, in Quantitative Methods for Health Research: A Practical Interactive Guide to Epidemiology and Statistics.* Chichester: John Wiley & Sons.

Crano, W.D., Brewer, M.B. (2002). *Principles and methods of social research* (ediţia a II-a). Mahwah, NJ: Lawrence Erlbaum Associates.

Cronbach, L. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, pp. 671-684.

Crosby, D.A., Dawsett, C.J., Gennetian, L.A., Huston, A.C. (2010). A tale of two methods: Comparing regression and instrumental variables estimates of the effects of preschool child care type on the subsequent externalizing behavior of children in low-income families. *Developmental Psychology*, 46(5), pp. 1030-1048.

Culic, I. (2004). *Metode avansate în cercetarea socială*. Iaşi: Editura Polirom.

Datculesu, P. (2006). *Cercetarea de marketing, Cum pătrunzi în mintea consumatorului, cum măsori informaţia şi cum analizezi informaţia.* Bucureşti: Editura Brandbuilders Group.

David, D. (2006). *Metodologia cercetării clinice. Fundamente.* Iaşi: Editura Polirom.

Duncan, T.E., Duncan, S.C., Strycker, L.A. (2006). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications: Revised Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Elmes, D.G., Kantowitz, B.H., Roediger, H.L. (2012). *Research methods in psychology.* Monterey, CA: Wadsworth.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, pp. 581-586.

Grimes, D.A., Schulz, K.F. (2002). Bias and causal associations in observational research. *Lancet*, 359, pp. 248-252.

Hartley, J. (2004). Case study research. În C. Cassell şi G. Symon (eds.), *Essential guide to qualitative methods in organizational research* (pp. 323-333). Londra: Sage.

Holbrook, A.L., Krosnick, J.A., Pfent, A. (2007). The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. În J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. d. Leeuw, L. Japec, P.J. Lavrakas, M.W. Link şi R.L. Sangster (eds.), *Advances in Telephone Survey Methodology* (pp. 499-528). Hoboken, NJ: John Wiley & Sons.

Katz, M.H. (2010). *Evaluating Clinical and Public Health Interventions. A Practical Guide to Study Design and Statistics*. New York, NY: Cambridge University Press.

Kazdin, A.E.(2010). *Single-case research designs: Methods for clinical and applied settings* (ediţia a II-a). New York: Oxford University Press.

Kemppainen, L., Veijola, J., Jokelainen, J., Hartikainen, A.L., Järvelin, M.R., Jones, P., Croudace, T., Isohanni, M. (2001). Birth order and risk for schizophrenia: a 31-year follow-up of the Northern Finland 1966 Birth Cohort. *Acta Psychiatrica Scandinavica*, 104(2), pp. 148-152.

Labăr, A.V. (2008). *SPSS pentru ştiinţele educaţiei. Metodologia analizei datelor în cercetarea pedagogică.* Iaşi: Editura Polirom.

Lee, J.J. (2012). Correlation and causation in the study of personality. *European Journal of Personality*, 26, pp. 372-390.

MacKinnon, D.P. (2008). *Introduction to Statistical Mediation Analysis.* Mahwah, NJ: Erlbaum.

Mitchell, M.L., Jolley, J.M. (2006). *Research design explained* (ediţia a VI-a). Belmont, CA: Wadsworth.

Neuman, W.L. (1997). *Social research methods: Qualitative and quantitative approaches* (ediţia a III-a). Boston: Allyn and Bacon.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (ediţia a II-a). New York: Cambridge University Press.

Petrie, A., Sabin, C. (2009). *Medical Statistics at a Glance* (ediţia a III-a). Oxford: Wiley-Blackwell.

Pintea S., Moldovan, R. (2009). The receiver-operating characteristic (ROC) analysis: Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies*, 9(1), pp. 49-66.

Popa, M. (2008). *Statistica pentru psihologie. Teorie şi aplicaţii SPSS*. Iaşi: Editura Polirom

Rotariu, T., Iluţ, P. (2006). *Ancheta sociologică şi sondajul de opinie. Teorie şi practică*. Iaşi: Editura Polirom.

Sava, F.A. (2004/2011). *Analiza datelor în cercetarea psihologică.* Cluj-Napoca: Editura ASCR.

Sava, F.A. (2009). Maladaptive schemas, irrational beliefs, and their relationship with the five factor personality model. *Journal of Cognitive and Behavioral Psychotherapies*, 9(2), pp. 135-147.

Sava, F.A., Pătruţ, L.C. (2010). *Probele Flanker AC, Flanker AD şi PON.* Timişoara: Editura ArtPress.

Sava, F.A., Popa, R.I. (2011). Personality types based on the Big Five model. A cluster analysis over the Romanian population. *Cognition, Brain, Behavior. An Interdisciplinary Journal*, 15, pp. 359-384.

Singer J.D., Willett, J.B. (2003). *Applied Longitudinal Data Analysis. Modeling change and event occurrence.* New York: Oxford University Press.

Stock, J.H., Wright, J.H., Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, 20(4), pp. 518-529.

Sulea, C., Vîrgă, D., Maricuţoiu, L.P., Schaufeli, W., Zaborilă-Dumitru, C., Sava, F.A. (2012). Work engagement as mediator between job characteristics and positive and negative extra-role behaviors. *Career Development International*, 17(3), pp. 188-207.

Ullebø, A.K., Posserud, M.B., Heiervang, E., Gillberg, C., Obel, C. (2011). Screening for the attention deficit hyperactivity disorder phenotype using the strength and difficulties questionnaire. *European Child & Adolescent Psychiatry*, 20(9), pp. 451-458.

Vîrgă, D. (2007). *Psihologie experimentală – de la teorie la practică*. Timişoara: Editura Universităţii de Vest.

Wegener, D.T., Fabrigar, L.R. (2000). Analysis and design for nonexperimental data: addressing causal and noncausal hypotheses. În H.T. Reis şi C.M. Judd (eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 412-450). New York, NY: Cambridge University Press.

World Health Organization (2001). *The world health report. Mental health: new understanding, new hope.* Disponibil la http://www.who.int/whr/2001/en/whr01_en.pdf.

1. Este vorba aici de absenţa controlului care ar fi fost exercitat prin manipularea variabilei cu potenţial cauzal. Alte tipuri de control (controlul în selecţia participanţilor sau controlul statistic) pot fi prezente în studiile non-experimentale.

2. Analizele predictive derulate în cadrul unor studii non-experimentale îşi păstrează o valenţă pur descriptivă. În schimb, analizele predictive efectuate în context experimental pot constitui o bază relevantă de discuţie în termenii unei relaţii de tip cauză-efect.

3. Nu toate studiile de caz au o asemenea structură comprehensivă. Unele studii de caz vizează doar aspecte specifice, cum ar fi implementarea unei tehnici cunoscute într-un context nou (de exemplu,

tehnici de hipnoză în coaching-ul organizaţional; descrierea factorilor antecedenţi declanşării unei tulburări etc.). Structura studiilor de caz variază şi în funcţie de perspectiva metodologică utilizată. Un exemplu de studiu de caz care apelează la abordare narativă (metodologie calitativă) poate fi articolul semnat de unul dintre cei mai talentaţi jurnalişti români contemporani, Vlad Mixich (2012), fiind disponibil la http://www.hotnews.ro/stiri-esential-13459189-semnatura-viata-moarte.htm.

4. Rotariu şi Iluţ (2006) definesc sondajele de opinie ca o specie populară a anchetei sociologice, centrată pe aspectul opinional, subiectiv al realităţii. În sociologie este cunoscută distincţia dintre opinii (cele mai instabile), atitudini (relativ stabile) şi valori (cele mai stabile), toate fiind aspecte subiective ale realităţii. În acest context, opiniile reprezintă reacţii verbale, relativ fluctuante, ce pornesc de la anumite atitudini, în timp ce acestea din urmă, la rândul lor, derivă din valori. Desigur, ceea ce sociologii au denumit anchete sociologice, economiştii le-ar putea eticheta ca studii de piaţă, iar specialiştii în medicina ca studii epidemiologice.

5. În cazul eşantionării stratificate se selectează deseori un număr egal de persoane din fiecare strat pentru a permite comparaţii între categoriile unui strat. De exemplu, într-un sondaj au fost aplicate câte 300 de chestionare în fiecare din cele patru filiale ale unei companii, pentru a face comparaţii între filiale, deşi filialele A şi B au un număr dublu de angajaţi faţă de C şi D. În asemenea cazuri este important să fie ponderate răspunsurile lor atunci când sunt analizate răspunsurile la nivel global (al întregii companii). De exemplu, dacă procentul de femei din filialele A, B, C şi D este de 20%, 30%, 50% şi 60%, procentul real de femei de la nivelul întregii companii este de 35% – obţinut ca medie aritmetică ponderată ((2×0,20)+(2×0,30)+(1×0,50)+(1×0,60))/6 – şi nu 40%, cât ar fi fost o medie aritmetică simplă. Ideea ponderării rezultatelor poate fi aplicată şi în cazul altor tipuri de eşantionări, dacă cazurile selecţionate nu respectă proporţia existentă la nivelul populaţiei din care a fost extras eşantionul, cu privire la anumite variabile importante.

6. Reprezentativitatea unui eşantion este un aspect indirect legat de volumul eşantionului. În principiu, cu cât un eşantion are un volum mai mare, există o probabilitate mai mare ca el să fie unul reprezentativ pentru populaţia studiată. Însă reprezentativitatea vizează eterogenitatea (diversitatea) indivizilor cuprinşi în studiu. Un eşantion format din 100 de persoane de ambele sexe, de vârste diferite, cu nivel educaţional şi localităţi de rezidenţă variate este mai reprezentativ pentru a descrie populaţia adultă a României decât un eşantion de 1.000 de persoane, toate de gen feminin, provenite din mediul rural, cu studii medii şi vârsta cuprinsă între 45 şi 65 de ani.

7. În Rotariu şi Iluţ (2006) veţi găsi mai multe detalii, inclusiv o tipologie comprehensivă despre tipurile de întrebări şi principiile lor de construire.

8. O variantă a studiilor panel clasice sunt studiile panel pivotante (*revolving panel study*), în care cea mai mare parte a participanţilor este aceeaşi, însă se permite să se înlocuiască un procent redus din participanţii vechi cu alţii noi, pentru a acoperi pierderile inerente de participanţi.

9. Acest tip de design este întâlnit şi în domeniul ştiinţelor sociale ca design pe cohorte secvenţiale (cohort-sequential design). De exemplu, cohorta de născuţi din perioada 1976-1980 este comparată cu cea care cuprinde pe cei născuţi în perioada 1986-1990, utilizând aceiaşi subiecţi, măsuraţi repetat în 2003, în 2008 şi în 2013 cu privire la atitudinea faţă de avort. Interpretarea rezultatelor din asemenea studii este foarte interesantă. Dacă rezultatele nu se replică (sunt diferenţe între cele două cohorte) se poate pune pe seama specificului diferit (caracteristicii) al unei cohorte. Dacă rezultatul

se replică la un anumit moment în timp (de exemplu, la evaluarea din 2008), deşi indivizii din cohorte au vârste diferite, am putea pune rezultatul pe seama unui eveniment apărut în societate, care îi afectează la fel, indiferent de timp. În sfârşit, dacă rezultatul se replică pe cohorte la intervale de timp egale cu distanţa dintre cohortă (atitudinea celor născuţi între 1976-1980 despre avort în 2003 este aceeaşi cu cea a celor născuţi în perioada 1986-1990 măsurată în 2013), am putea pune efectul pe seama vârstei (unei anumite etape de dezvoltare, cum ar fi opinia adolescenţilor la 15 ani despre avort).

10. Fiecare strategie longitudinală poate fi implementată într-o manieră prospectivă (din prezent spre viitor) sau într-o manieră retrospectivă (dinspre trecut spre prezent, dacă există acces la date din trecut pe care să ne putem încrede).

11. Pentru mai multe detalii despre variabilele confundate şi când trebuie acestea controlate, vezi caseta 5.4.

12. Controlul statistic al variabilelor confundate constituie doar o premisă, nu şi o garanţie pentru a face afirmaţii de tip cauzal (vezi problema scorurilor de propensitate discutate în secţiunea 5.5).

13. Studiile longitudinale, deşi permit adresarea unei probleme de bază în stabilirea unei relaţii cauză-efect, aceea de a arăta că modificările variabilei-cauză le precedă temporal pe cele ale variabilei-efect (întâi te apuci de fumat şi apoi apare cancerul pulmonar), nu sunt infailibile în acest sens. De exemplu, să ne imaginăm ipotetic că există o genă responsabilă pentru nivelul apropierii de oameni, care tinde să conducă atât la un ataşament timpuriu insecurizant, cât şi la probleme de relaţionare în viaţa de cuplu ca adult. În absenţa informaţiilor despre o asemenea genă, există riscul ca, derulând un studiu longitudinal, să observăm că cei cu ataşament insecurizant tind să dezvolte relaţii problematice de cuplu ca adult şi să punem acest lucru pe seama stilului de ataşament insecurizant format în copilărie (pe care eventual să încercăm să-l restructurăm în cadrul unui proces psihoterapeutic). În realitate, dacă o asemenea genă ar exista, ar însemna că ambele probleme sunt doar efecte ale unei anumite structuri genetice, ce se diferenţiază prin perioade diferite de apariţie – ataşamentul insecurizant din copilărie precedând temporal problemele ulterioare de relaţionare în cuplu.

14. Cei cu experienţă în domeniul studiilor randomizate ştiu cât de dificil poate fi să nu existe o rată mare de pierdere a participanţilor dacă perioada de urmărire durează 2-3 luni. Efortul este cu atât mai mare, cu cât este vorba de perioade de urmărire de ani sau chiar zeci de ani.

15. Este vorba de un anumit tip de prevalenţă, cea raportată la un anumit moment din timp (*point prevalence*). Se poate calcula şi o prevalenţă pentru un interval de timp (prevalenţa anuală, prevalenţa pe durata întregii vieţi), situaţie care poate fi abordată mai eficient în cadrul unor studii longitudinale, deşi nu este necesar apelul la testarea aceloraşi subiecţi.

16. La numitorul formulei de calcul a prevalenţei pentru un anumit interval de timp se ţine seama de populaţia existentă la jumătatea perioadei analizate (în cazul de faţă, populaţia rămasă la 01.12.2012). Cum la acel moment nu a existat o asemenea evaluare, se poate estima fie o medie a celor două valori colectate (1.000, respectiv 920), fie se poate lua în calcul ultima valoare colectată care precedă jumătatea intervalului avut în vedere, în cazul dat, valoarea 1.000 existentă la debutul studiului. Din valoarea iniţială nu se exclud doar pierderile, ci şi ieşirile naturale precum moartea unor persoane (ei nu mai fac parte din populaţia respectivă) sau situaţii specifice precum (migrarea sau ieşirea din grupul de interes – de exemplu, nu mai sunt în cadrul vârstei-ţintă cuprinse între 14 şi 18 ani).

17. În cazul ratei incidenţei, formula se complică pentru numitor. Dacă la numărător este vorba de numărul de cazuri noi apărute în perioada de urmărire analizată, la numitor se calculează suma perioadelor de timp până la apariţia cazului, în cazul persoanelor cu risc de a apărea acest lucru. În exemplul oferit, persoane cu risc sunt cele 940 care nu consumau la momentul iniţial droguri. În cazul fiecărei persoane care rămâne şi la a doua măsurare în categoria neconsumatorilor de droguri se adaugă un an de zile (840 × 1 an = 840 de ani). În cazul persoanelor care între timp au început să consume se adaugă perioada de timp în care nu au consumat (de exemplu, dacă un individ a reuşit să rămână 8 luni din cele 12 cuprinse între două evaluări fără să consume droguri, ar însemna 0,66 ani). În exemplul oferit, deşi nu sunt oferite valorile individuale pentru fiecare dintre cei 30 de noi consumatori, se specifică faptul că în medie consumă de 6 luni (0,5 ani × 30 = 15 ani). În sfârşit, în exemplu au fost excluşi din calcul cei pierduţi, pentru că nu există două măsurători pentru aceştia. Dacă ar fi existat o evaluare intermediară după 6 luni, în care din cei 70 pierduţi la final ar fi existat rezultate intermediare pentru 50 dintre ei (toate cazuri fără consum de droguri), s-ar fi adăugat la totalul de 855 de ani, încă 25,5 (50 × 0,5 ani). În pofida formulei complicate, rata incidenţei are avantajul de a se raporta la un interval de timp standard (cum ar fi număr de cazuri noi/anual), în timp ce incidenţa cumulativă este dependentă de perioada variabilă dintre două testări (de exemplu, 6 luni etc.).

18. În cazul riscului relativ (o denumire alternativă întâlnită este cea de raportul şanselor – *hazard ratio* sau *odd ratio*), o valoare mai mare decât 1 indică o asociere pozitivă între predictor şi apariţia efectului, iar un rezultat subunitar indică o asociere negativă. În cazul riscului atribuibil, un rezultat pozitiv indică existenţa unui factor de risc, în timp ce un rezultat negativ ar fi indiciul un factor protectiv (dacă efectul urmărit este unul nedorit).

19. Ar fi contextul potrivit pentru a face *mea culpa*. În 2002 am publicat primul meu articol indexat într-o revistă ISI, testând prin analiză de cale o serie de antecedente şi consecinţe ale unui comportament inadecvat al profesorului la clasă. L-am denumit pretenţios *Causes and effects of teacher conflict-inducing attitudes towards pupils: a path analysis model*, deşi studierea unei relaţii de tip cauză-efect ar fi presupus apelul la un studiu experimental, iar acolo unde nu este posibil, la un studiu cvasiexperimental sau la unul non-experimental longitudinal (pentru a observa că modificările în variabila-cauză le precedă pe cele observate la nivelul variabilei efect). Problema limitelor existente în adresarea unor relaţii de tip cauză-efect prin studii corelaţionale transversale este prezentată în secţiunea 5.4.2. Îmi amintesc, pe de altă parte, că a fost primul studiu psihologic din România care a folosit analiza de cale. La cea mai importantă conferinţă naţională de la acea vreme, organizatorii au modificat titlul lucrării, sintagma „analiză de cale" fiind înlocuită cu cea de „analiză de caz", deşi era evident din rezumat că nu era vorba despre un studiu de caz.

20. Bollen (1989 *apud* Wegener şi Fabrigar, 2000) denumeşte aceste trei condiţii: (a) existenţa asocierii; (b) direcţia asocierii şi (c) izolarea asocierii.

21. Un lucru care ar trebui precizat pentru toată lumea este distincţia dintre studiile corelaţionale, ca design non-experimental, şi tehnicile corelaţionale, ca metode de analiză statistică. Tehnicile statistice, în sinea lor, nu se diferenţiază în tehnici care să permită inferenţe cauzale şi tehnici care nu permit acest lucru. De exemplu, datele dintr-un studiu experimental pot fi analizate prin tehnici ANOVA, cum la fel de bine pot fi analizate şi prin regresii în care predictorii referitori la condiţiile

experimentale au fost codaţi *dummy*. Aşadar, e incorect să se afirme că apelul la tehnica ANOVA permite inferenţe cauzale, în timp ce regresiile nu permit acest lucru. Ceea ce contează este designul utilizat şi strategia de analiză a datelor. De exemplu, în cazul regresiei multiliniare ierarhice, elementul care face plauzibile anumite interpretări cauzale mai mult decât studiile corelaţionale clasice este faptul că analiza relaţiei dintre potenţialul VI şi VD se face în contextul controlului statistic al impactului altor factori relevanţi asupra VD. De aceea, aş considera eticheta oferită acestei secţiuni (regresie ierarhică) ca fiind o denumire generică, aplicată în domenii multiple (regresii multiliniare, regresii logistice, modelare multinivelară), dar care au ca element central comun, analiza în paşi, pentru a controla efectul unor variabile relevante.

22. Asendorpf (2012) subliniază rolul acestor variabile în interpretarea tehnică a unor rezultate statistice. Astfel, în studii bazate pe performanţa şcolară, deşi atât IQ-ul, cât şi conştiinciozitatea sunt identificaţi adesea ca predictori ai unei performanţe superioare, se pot observa adesea corelaţii negative între IQ şi conştiinciozitate. Această corelaţie negativă este o consecinţă a efectului pe care variabila de ciocnire, în acest caz performanţa şcolară, îl joacă în dinamica celor doi factori (predictori) care contribuie împreună la o performanţă şcolară mai bună. Prin urmare, o corelaţie între două variabile poate avea multiple surse: a. existenţa unei relaţii cauzale; b. dependenţa comună a variabilelor corelate de o cauză comună (corelaţii false); c. contribuţia comună a două variabile cauză asupra unui efect comun (corelaţie falsă).

23. De exemplu, în literatura de specialitate de la noi, Sava (2004/2011) prezintă în capitolul 6, în detaliu, modul de implementare a celor trei tipuri majore de relaţii: directe, mediate şi moderate. Ele ar putea fi completate de alte situaţii complexe, cum ar fi: (1) testarea efectului de mediere moderată; respectiv (2) a celui de moderare mediată (MacKinnon, 2008; Wegener şi Fabrigar, 2000). În primul caz, se presupune: (a) identificarea unui termen de interacţiune dintre moderator şi mediator semnificativ statistic, (b) acompaniat de o reducere a efectului interacţiunii dintre predictor şi moderator, (c) după ce, în prealabil, s-au controlat statistic efectele determinate de interacţiunea predictor × moderator şi cele determinate de efectele principale (predictor, respectiv variabilă moderatoare). Interpretarea dată ar fi aceea că impactul variabilei cu rol mediator între VI şi VD este diferit în funcţie de nivelul variabilei moderatoare. În cel de-al doilea caz, variabila moderatoare nu exercită rolul său direct asupra relaţiei dintre VI şi VD, ci indirect prin intermediul unei variabile mediatoare. În acest caz se urmează procedura tipică pentru analiza de mediere prin regresie (vezi detalii în Sava, 2004/2011 sau Baron şi Kenny, 1986), doar că în locul variabilei predictor (X) se va utiliza termenul de interacţiune între predictor şi variabila moderatoare, desigur, dacă în prealabil a existat un asemenea efect de interacţiune semnificativ statistic.

24. O cutumă este de a accepta ca variabile instrumentale numai pe cele care, într-o ecuaţie de regresie în care variabilele instrumentale prezic nivelul variabilei-cauză se obţine un F semnificativ statistic, cu o valoare cel puţin egală cu 10 (Stock, Wright şi Yogo, 2002).

25. Pentru situaţiile care nu implică alte variabile confundate, Z trebuie să coreleze cu X, respectiv Z trebuie să nu coreleze cu valoarea reziduală rămasă în urma prezicerii lui Y pe baza lui X.

26. Denumirea provine din domeniul studiilor medicale, în care sunt comparate două grupe, cum ar fi grupul cazurilor de interes (de exemplu, fumători) cu cel de control (nefumători, cu caracteristici

sociodemografice similare cu cele ale fumătorilor) în privinţa unei anumite variabile precum capacitatea de rezistenţă fizică la efort.

27. Cu o singură condiţie: să fie vorba de grupe suficient de mari ca număr de subiecţi. Randomizarea în grupe mici este un proces failibil, de aceea în cazul lor vorbim de grupe potenţial echilibrate.

28. Cei familiarizaţi cu elemente de algebră au intuit, probabil, că din punct de vedere matematic: LR+ = sensibilitatea/(1 – specificitatea), iar LR– = (1 – sensibilitatea)/specificitatea.

29. Formulele de calcul pentru valorile ponderate sunt: LR+ ponderat = LR+ × [prevalenţa/(1 – prevalenţa)], iar LR – ponderat = LR– × [prevalenţa/(1 – prevalenţa)]. De exemplu, în cazul unei prevalenţa a ADHD de 10% (0,10), avem o valoare LR+ ponderată = 6,46 respectiv o valoare LR– ponderată = 0,02. Prin ponderarea aplicată, interpretarea rezultatelor se schimbă: (a) există de peste 6 ori mai multe şanse ca un test screening pozitiv să fie întâlnit la persoane cu ADHD decât la cele fără ADHD; (b) există de 50 de ori mai puţine şanse ca un test screening negativ să fie întâlnit la persoane cu ADHD decât la cele fără ADHD. Pentru a transforma apoi asemenea valori ajustate în probabilitatea ca o persoană cu test pozitiv, respectiv negativ să aibă într-adevăr tulburarea respectivă, valorile ponderate LR se transformă în probabilităţi posttest. În cazul LR+ ponderat vom obţine o probabilitate de 0,865 sau 86,5% – 6,46/(1 + 6,46) – ca o persoană cu un rezultat pozitiv la testul de screening să aibă ADHD (faţă de o probabilitate de doar 0,10 – 10% – în absenţa testului de screening, dat fiind că doar 10% din populaţie are ADHD). În cazul LR– vom obţine o probabilitate de doar 0,019 sau 1,9% – 0,02/(1 + 0,02) – ca o persoană cu un rezultat negativ la testul de screening să aibă ADHD, în contextul în care tulburarea are o prevalenţă de 10%.

30. Veţi găsi pe internet mai multe aplicaţii care fac toate aceste calcule pentru dumneavoastră. O asemenea aplicaţie este disponibilă gratuit la http://www.vassarstats.net.

31. Spre ilustrare, dacă veţi consulta articolul original a lui Ullebø şi colaboratorii (2011) de unde a fost inspirat acest exemplu, veţi observa că prevalenţa ADHD a fost de 5%, iar valorile sensibilităţii şi specificităţii SDQ, versiunea pentru profesori, au fost de 0,84, respectiv 0,88 pentru un scor-limită pentru un test pozitiv ≥ 5; de 0,59 şi 0,96 pentru un scor-limită pentru un test pozitiv ≥ 7; şi de 0,33, respectiv 0,99 pentru un scor-limită al unui test pozitiv ≥ 9.

Capitolul 6

Studiile de evaluare economică a serviciilor psihologice

La o primă vedere, cititorii pot fi surprinşi de decizia de a rezerva un capitol întreg discutării fundamentelor evaluării economice într-o lucrare de psihologie, chiar dacă discuţia are loc în contextul furnizării de servicii de evaluare şi intervenţie psihologică. Pe parcursul capitolelor anterioare s-a accentuat ideea unei practici psihologice validate ştiinţific prin cercetare. Însă orice serviciu de natură psihologică implică un consum de resurse, implicit costuri. Prin urmare, serviciul trebuie nu doar să fie unul validat ştiinţific, ci şi unul eficient din perspectivă financiară. Piaţa serviciilor psihologice creează un mediu concurenţial. Prin urmare, de cele mai multe ori se pune problema nu numai în a stabili care serviciu este eficient, ci şi în a determina care dintre serviciile funcţionale are cel mai bun raport cost-eficienţă. Aşadar, miza unei asemenea evaluări economice are loc în contextul în care există cel puţin două intervenţii eficiente pentru a rezolva aceeaşi problemă, urmând a se stabili care dintre acestea este mai eficientă sub raport economic.

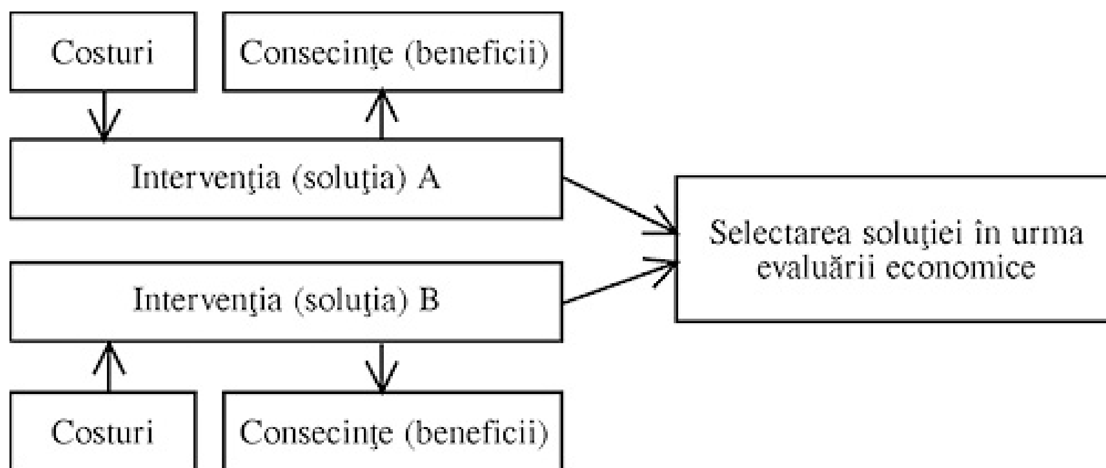


Figura 6.1. *Schema logică de bază a unei evaluări economice*

Acest tip de gândire domină atât mentalul beneficiarilor serviciilor de psihologie din domeniul organizaţional (decidenţii din companii fiind primii interesaţi să afle „cât costă serviciul propus?” sau „care este câştigul adus?”), cât şi starea de fapt întâlnită în domenii precum cel medical sau cel şcolar, confruntate cu bugete limitate. Din acest punct de vedere, cel mai avansat sistem decizional în baza evaluărilor economice este întâlnit în Marea Britanie (Taylor, 2009). În această ţară serviciile de sănătate sunt supuse unei evaluări economice, fiind finanţate doar acele intervenţii care se dovedesc a avea cel mai bun raport cost-eficienţă/cost-utilitate.

A ignora această tendinţă naturală elementară de a selecta acele servicii psihologice care maximizează efectele dorite în condiţiile minimizării costurilor înseamnă a ignora regulile simple ale unei pieţe libere sau ale unei societăţi ce dispune de resurse (bugete) limitate pentru a acoperi serviciile psihologice necesare populaţiei. O astfel de neglijare contribuie la marginalizarea serviciilor psihologiei (din sfera serviciilor de sănătate sau educaţie publică finanţate direct sau indirect de către stat) şi la o reducere a atractivităţii unor asemenea servicii în domeniul organizaţional (mediul privat). Spre ilustrare, oricâte mesaje emoţionale ar fi transmise către autorităţi privind necesitatea subvenţionării unor servicii de psihoterapie către populaţia generală, ele nu vor avea acelaşi impact asupra decidenţilor publici precum rezultatul obţinut în urma analizei efectuate de Cummings (1977, *apud* Yates, 1994). Conform acestei analize, pentru fiecare dolar cheltuit cu subvenţionarea unor şedinţe scurte de psihoterapie statul câştigă în medie 2,59 dolari prin eliminarea altor costuri (servicii asistenţiale) şi prin creşterea productivităţii clienţilor care au urmat şedinţele de psihoterapie.

Dacă asupra nevoii de realizare a unor asemenea analize există o unitate de opinii, îndeosebi din partea decidenţilor interesaţi să apeleze la servicii cu costuri cât mai reduse, dar cu eficienţă ridicată, există o diversitate de păreri cu privire la modul de realizare a unei asemenea analize. În secţiunea următoare vom introduce succint principalele tipuri de analize economice bazate pe raportul dintre resursele consumate şi valoarea rezultatelor obţinute. În celelalte secţiuni vom prezenta mai multe detalii despre fiecare dintre acestea, insistând asupra analizei cost-eficienţă şi

asupra analizei cost-utilitate datorită amplorii cu care sunt utilizate în evaluările economice ale serviciilor din sfera psihologiei şi a sănătăţii mentale.

6.1. Tipuri de evaluări economice şi criterii de selectare a lor

Toate evaluările economice, cu unele rare excepţii, au drept caracteristică comună faptul că măsoară atât *inputul* (întotdeauna sub forma unor costuri necesare implementării unui serviciu), cât şi *outputul* (beneficiile obţinute în urma implementării acelui serviciu). Evaluările diferă în funcţie de tipul de *output* valorizat, acestea fiind discutate la momentul prezentării fiecărui tip de analiză şi sintetizate în tabelul 6.1, adaptat după Haycox (2009).

6.1.1. Analiza minimizării costurilor

Acest tip minimalist de evaluare economică este cunoscut şi sub denumirea simplă de analiză de costuri deoarece ia în considerare doar costurile implicate, nu şi eficienţa unui serviciu psihologic. De aceea, decizia luată în acest caz este una simplă, fiind selectat serviciul cel mai ieftin. Din nefericire, mulţi oameni de decizie din sfera publică se rezumă la acest gen de analiză, abordarea fiind încurajată de o serie de reglementări legislative, cum ar fi alegerea ofertantului cel mai ieftin. Focalizarea în exclusivitate pe costuri, fără a lua în considerare diferenţele dintre servicii la nivel de beneficii, poate fi păguboasă pe termen mediu şi lung, ajungându-se adesea la costuri suplimentare ulterioare. Fără a absolutiza, adesea lucrurile care costă mai puţin au o durată de garanţie mai redusă sau conduc la efecte pozitive de intensitate mai scăzută în comparaţie cu alte soluţii mai scumpe. De aceea, acest tip de analiză este potrivită spre a fi implementată într-o singură situaţie, aceea în care intervenţiile comparate (supuse evaluării) au acelaşi nivel de eficienţă, fiind comparabile din perspectiva consecinţelor dezirabile pe termen mediu şi lung. Această condiţie este destul de greu de îndeplinit, deoarece ar

presupune existenţa prealabilă a unor studii randomizate controlate care să indice echivalenţa sau măcar non-inferioritatea formelor de intervenţie supuse comparării (vezi criteriul scopului comparaţiei din capitolul 4, tabelul 4.4). În absenţa unor asemenea dovezi ale echivalenţei consecinţelor rezultate din serviciile supuse comparării, rezultatele analizei pot conduce la decizii greşite din punct de vedere economic.

6.1.2. Analiza cost-eficienţă

Este cel mai întâlnit tip de analiză în domeniul serviciilor psihologice, fiind o analiză care ia în considerare atât costurile (exprimate în termeni financiari), cât şi beneficiile obţinute (exprimate în termeni nefinanciari). Acest tip de analiză poate fi efectuat pentru a compara eficienţa unor servicii de psihologie care vizează aceeaşi problemă şi care folosesc un numitor comun pentru a exprima beneficiile. De exemplu, se poate compara eficienţa unui program de coaching individual menit să îmbunătăţească nivelul de autoeficacitate al noilor angajaţi cu eficienţa unei intervenţii colective prin traininguri centrate pe acumularea de cunoştinţe tehnice, dacă ambele programe urmăresc acelaşi deziderat final – reducerea timpului de asistenţă oferită noilor angajaţi în vederea integrării în companie –, operaţionalizat prin costul per zi fără asistenţă oferită noului angajat. Un alt exemplu din domeniul clinic vizează serviciile menite să combată depresia. Pot fi comparate terapia cognitiv-comportamentală cu terapia familial-sistemică din perspectiva costului serviciului pentru a obţine o zi fără simptomatologie depresivă. Utilizând o asemenea analiză într-un studiu anterior am arătat, de exemplu, că pentru spaţiul economic al României este mai rentabil să se apeleze la psihoterapie cognitiv-comportamentală decât la medicaţie în combaterea depresiei majore, deoarece aceasta are un raport cost-eficienţă mai bun decât medicaţia standard utilizată – fluoxetina (cu denumirea comercială Prozac) (Sava, Yates, Lupu, Szentagotai şi David, 2009).

Beneficiile avute în considerare vizează problema investigată (costul unui an fără *burnout*/angajat; costul unei luni fără simptomatologie depresivă/pacient; costul anual al programului pentru reducerea fluctuaţiei de personal per angajat

păstrat în companie etc.). Analiza cost-eficienţă permite compararea adecvată a programelor care urmăresc acelaşi obiectiv major, însă ea este nepotrivită pentru a compara servicii care urmăresc obiective multiple (de exemplu, atât creşterea productivităţii angajaţilor, cât şi reducerea fluctuaţiei de personal). De asemenea, acest gen de analiză nu permite compararea eficienţei programelor dacă *outputul* urmărit este unul diferit (de exemplu, analiza cost-eficienţă nu poate răspunde la întrebarea tipică întâlnită într-un context de constrângeri bugetare – ar fi mai potrivit ca statul să sprijine un program de reducere a anxietăţii sociale sau un program de reducere a dependenţei de alcool?). Acest fapt se întâmplă deoarece nu se poate pune un semn de egalitate între dependenţa de alcool şi anxietatea socială din perspectiva implicaţiilor pentru calitatea şi durata vieţii unui individ.

6.1.3. Analiza cost-utilitate

Acest tip de analiză se apropie mult de ideea unei analize de tip cost-eficienţă, având, asemeni acesteia, costuri exprimate în termeni financiari şi efectele pozitive exprimate în termeni nefinanciari. Spre deosebire de analiza de tip cost-eficienţă, analiza cost-utilitate permite realizarea de comparaţii privind eficienţa economică a unor servicii destinate unor obiective diferite (de exemplu, este mai eficient să fie finanţat un serviciu de psihoterapie cognitiv-comportamentală pentru a reduce disfuncţiile erectile sau un serviciu de terapie comportamentală de tip ABA în cazul copiilor autişti?). Acest tip de analiză este întâlnit doar în domeniul larg al serviciilor de sănătate, fiind posibil prin introducerea la nivel de *output* a unui numitor comun, de obicei exprimat prin QALY – *Quality Adjusted Life Years*. Acest indicator ţine seama de utilitatea percepută a serviciilor comparate în funcţie de tulburarea urmărită. Analiza cost-utilitate, prin numitorul comun pe care-l oferă, permite compararea eficienţei economice a unei intervenţii focalizate pe renunţarea la fumat cu cea a unei program de consiliere pentru diabet sau cu cea a unei intervenţii chirurgicale de sudare a unei fracturi de femur etc. Logica analizei cost-utilitate poate fi transferată şi în domeniul organizaţional (ar trebui încurajat un program de stimulare a celor cu performanţă ridicată sau un program de reducere a ratei fluctuaţiei de personal?). Cu toate acestea, metodologia de lucru implică o

serie de decizii cu caracter relativ (subiectiv), fiind aduse o serie de critici de natură filosofică şi tehnică (de exemplu, cine ar fi cel mai îndreptăţit să acorde o valoare subiectivă – cunoscută tehnic ca utilitate: comunitatea, specialiştii în medicină, pacienţii cu problemele respective etc.).

6.1.4. Analiza cost-beneficiu

Analiza cost-beneficiu reprezintă modelul ideal de analiză economică, în cadrul căreia atât costurile serviciului oferit, cât şi beneficiile rezultate din această stare de fapt sunt exprimate în termeni financiari. Acest lucru este mult mai uşor de obţinut în domeniul organizaţional (de exemplu, pot fi estimate costurile necesare înlocuirii unui om plecat din companie într-un studiu despre fluctuaţia de personal). Este însă mult mai dificil de calculat acest lucru în domeniul sănătăţii (care ar fi valoarea exprimată în termeni financiari a unui caz însănătoşit de anxietate socială?). De aceea, acest tip de analiză este aproape absentă din sfera psihologiei clinice şi a psihoterapiei, fiind ceva mai prezentă în zona serviciilor organizaţionale ce permit urmărirea unor obiective financiar tangibile (de exemplu, costul unui serviciu de motivare a angajaţilor conduce la o creştere a productivităţii medii în muncă cu 15%).

Tabelul 6.1. *Grila de selectare a celui mai potrivit tip de analiză în funcţie de poziţionarea faţă de beneficiile serviciului psihologic oferit (adaptare după Haycox, 2009)*

Tipul de evaluare economică	***Operaţionalizarea avantajelor rezultate din serviciul oferit***	***Interpretare/decizie luată***
Analiza costului minim	Se calculează doar costurile implicate, fiind utilă doar în contexte în care există dovezi privind echivalenţa efectelor intervenţiilor comparate.	Dacă există date privind echivalenţa eficienţei intervenţiilor, se alege soluţia cea mai ieftină.
Analiza cost-eficienţă (CEA)	Se calculează atât costurile financiare, cât şi avantajele exprimate într-o manieră nefinanciară, dar care au o semnificaţie naturală, cum ar fi schimbarea în bine a stării de sănătate a unui pacient. CEA permite doar compararea de programe care vizează acelaşi obiectiv (de exemplu, cost anual/persoană fără depresie).	Se calculează pentru fiecare program evaluat raportul dintre costurile implicate şi beneficiul obţinut, după care se pot calcula diferenţe incrementale, care să indice superioritatea unui program în dauna altuia din perspectiva raportului cost-eficienţă.

Analiza cost-utilitate	Se calculează atât costurile, cât şi avantajele rezultate, acestea din urmă fiind exprimate tot într-o manieră nefinanciară. Eficienţa este transpusă în indicatori standardizaţi (de exemplu, QALY), ceea ce permite şi comparaţii între programe care vizează obiective diferite (de exemplu, reducerea simptomatologiei legate de autism *vs* reducerea depresiei).	Se bazează pe o interpretare similară cu cea prezentă în cazul analizei de tip cost-eficienţă, cu precizarea că este vorba de costul pentru fiecare QALY.
Analiza cost-beneficiu	Atât costurile, cât şi beneficiile sunt exprimate în termeni financiari. Poate fi implementată doar dacă beneficiile pot fi transformate relativ direct în câştiguri financiare. Acest aspect este mai uşor îndeplinit în cazul serviciilor psihologice din domeniul organizaţional, îndeosebi în cele în care performanţa este măsurată financiar.	Intervenţia selectată este aceea care asigură cel mai mare profit (venituri estimate – costuri estimate). Pot fi implementate doar intervenţii cu o valoare pozitivă a acestui raport (veniturile sunt mai mari decât costurile implicate – *return of investment/ROI*).

6.2. Analiza cost-eficienţă

Pentru a înţelege cum se realizează o analiză cost-eficienţă, vom diviza această secţiune în unităţile componente: (a) estimarea costurilor; (b) estimarea eficienţei; (c) raportul cost-eficienţă şi (d) analiza de sensibilitate a acestui raport. Pentru a ilustra mai bine aceste aspecte, vom apela la un exemplu din domeniul clinic, pornind de la studiul efectuat pe un lot de pacienţi suferinzi de depresie majoră (Sava *et al*., 2009). De asemenea, pentru a întări cele ilustrate, vom oferi o ilustrare a analizei cost-eficienţă în domeniul organizaţional.

6.2.1. Estimarea costurilor

În estimarea costurilor se au în vedere două alternative diferite de estimare: costurile financiare pe care le implică intervenţia în sine şi costurile la nivel societal (economice). De exemplu, în cadrul unui program de consiliere individuală pentru a reduce riscul delincvenţei, costurile financiare s-ar limita la plata resurselor implicate (costurile orelor de consiliere, ale spaţiului utilizat pentru acestea, costurile de transport implicate, alte costuri indirecte necesare). În schimb, o abordare economică (la nivel societal) a costurilor ar adăuga la cele menţionate deja

şi costurile de oportunitate (Haycox, 2009) care ar include costurile pacientului (timpul petrecut de acesta în şedinţe de consiliere, care înseamnă pierderi de bani cu orele libere luate dacă el este angajat); timpul neproductiv petrecut de familie pentru sprijinirea celui consiliat; banii pe care statul îi salvează (reprezentând costurile de încarcerare, judecată etc.), dacă individul consiliat este recuperat şi nu mai comite acte infracţionale.

La ora actuală tinde să predomine o perspectivă favorabilă estimării costurilor societale (Fals-Stewart, Yates şi Klostermann, 2005) într-o versiune relativ moderată, aceea de a cuprinde acele costuri care pot fi estimate relativ uşor. Cea mai potrivită definiţie a costurilor implicate este aceea de valoare financiară a resurselor consumate ca urmare a implementării serviciului psihologic sau care au fost pierdute temporar ca urmare a tulburării respective (de exemplu, absenteism de la locul de muncă, reducerea programului de lucru etc.).

În privinţa costurilor legate de implementarea serviciilor de psihologie, acestea ar trebui să includă: (a) costurile legate de timpul necesar implementării serviciilor (de exemplu, costul per oră al unei şedinţe de psihoterapie; timpul alocat de client participării la aceste şedinţe, incluzând timpul de transport şi referirile la timpul persoanelor însoţitoare, acolo unde este cazul); (b) costurile legate de transportul spre şi de la locaţia unde se oferă intervenţia; (c) costurile legate de resursele materiale şi echipamentele necesare (probe de urină pentru programele antidrog, teste psihologice, de spitalizare etc.); (d) costuri de închiriere (legate de spaţiul unde se oferă intervenţia, dacă este cazul); (e) costuri indirecte, de regie. La aceste categorii de costuri se pot adăuga cele societale, cum ar fi costurile reprezentate de pierderile financiare ca urmare a instalării handicapului temporar (de exemplu, concedii medicale fără plată etc.) (Fals-Stewart, Yates şi Klostermann, 2005; Yates, 1999). Pentru fiecare dintre aceste costuri, se recomandă utilizarea unor tabele matriciale pentru a estima cât mai corect costurile. Aceste tabele sunt utile prin: (a) împărţirea serviciului în secvenţe; (b) stabilirea tipului de resurse implicate pentru fiecare secvenţă; (c) stabilirea numărului de unităţi de resurse necesare; (d) estimarea costului per unitate (la nivelul preţului mediu al pieţei); (e) obţinerea totalurilor marginale şi a sumei totale cu cheltuielile implicate.

Pentru a simplifica acest aspect, vom prezenta un tabel cu estimarea costurilor per pacient prezentat în Sava şi colaboratorii (2009). În cadrul acestuia se observă analiza consumului de resurse pe tipuri majore de activităţi implicate, divizat în cheltuieli financiare (pentru oferirea serviciilor respective) şi cheltuieli suplimentare ale clientului (pentru a oferi o perspectivă societală estimării acestor costuri). În studiul respectiv se urmărea realizarea unei analize cost-eficienţă în combaterea depresiei majore între tratamentul medicamentos prin fluoxetină (Prozac) şi două forme de psihoterapie de sorginte cognitiv-comportamentală, terapia cognitivă dezvoltată de Beck (CT), respectiv terapia raţional-emotivă dezvoltată de Ellis (REBT). Programul prevedea 20 de şedinţe de psihoterapie a câte 50 de minute fiecare, respectiv 20 de întâlniri informative/de monitorizare a efectului medicaţiei cu durata între 15 şi 20 de minute fiecare.

Tabelul 6.2. *Tabel centralizator de calcul estimativ al costurilor utilizând un model matricial Proceduri (secvenţe) × Resurse*

			Unităţi medii consumate per client			Total Cost per Client		
Tipul de resursă utilizată	Unitate de măsură	Preţ unitar	REBT	CT	Fluoxetină	REBT	CT	Fluoxetină
Manoperă/timp (tarif orar)								
Psihoterapeut	1 h	6,59 $	20,98 h	20,41 h	0 h	138,25 $	134,49 $	0 $
Psihiatru	1 h	8,09 $	0 h	0 h	6,95 h	0 $	0 $	56,25 $
Psiholog clinician	1 h	4,56 $	0,33 h	0,33 h	0,33 h	1,51 $	1,51 $	1,51 $
Asistent de cercetare	1 h	3,55 $	0,30 h	0,30 h	0,29 h	1,06 $	1,06 $	1,03 $
Cheltuieli cu spaţiu + utilităţi	1 m²/1 h	0,16 $	274,74 m²/h	267,64 m²/h	102,95 m²/h	43,52 $	42,39 $	16,31 $
Consumabile	1 FR/test/pers.	1,03 $	1,91 set	1,89 set	1,86 set	1,97 $	1,95 $	1,92 $
Licenţă teste	Manual/client	0,36 $	1 manual	1 manual	1 manual	0,36 $	0,36 $	0,36 $
Medicaţie	1 doză = 20mg	1,27 $	0 doze	0 doze	159,65 doze	0 $	0 $	202,31 $
Medicaţie de continuare	1 doză = 20mg	1,27 $	0 doze	0 doze	156,46 doze	0 $	0 $	198,27 $
Totalul cheltuielilor directe						*186,66*	*181,76*	*477,95* $

						$	$	
Regie (50% din totalul costurilor directe)						*93,33* $	*90,88* $	*238,97* $
Totalul costului serviciului						***279,99*** $	***272,64*** $	***716,92*** $
Timpul clientului	1 h	6,13 $	32,10 h	31,25 h	18,00 h	196,78 $	191,50 $	110,35 $
Costurile de transport	1 zi (drum dus-întors)	1,90 $	21,98/1 zi	21,41/1 zi	21,02/1 zi	41,78 $	40,70 $	39,95 $
Totalul costului de oportunitate						***238,56*** $	***232,20*** $	***150,30*** $
COSTUL TOTAL PE PACIENT	(Cost serviciu + Cost de oportunitate)/pacient					**518.55** $	**504.84** $	**867.22** $
COSTUL INTERVENŢIEI	= Nr. pacienţi × Costul total pe pacient					**29.557** $	**28.271** $	**49.431** $ **1.373,09** $
COSTUL TOTAL/PACIENT RECUPERAT	= COSTUL INTERVENŢIEI/Nr. pacienţi recuperaţi					**703,75** $	**724,90** $	

Explicitarea valorilor trecute în tabelul 6.2 ar necesita un spaţiu editorial generos. Detalii suplimentare pot fi consultate în Sava şi colaboratorii (2009). Aici ne vom rezuma la a prezenta aleatoriu câteva dintre cele mai importante costuri estimate. De exemplu, în coloana a patra se specifică o medie de 20,98 ore de psihoterapie de tip REBT urmate de un pacient din cele 23 maximum posibile (20 de ore din cadrul şedinţelor regulate, fiecare constând în 50 de minute de discuţii şi 10 minute antemergătoare de pregătire din partea terapeutului, la care se adaugă până la un maximum de 3 ore în cadrul şedinţelor suplimentare puse la dispoziţie în perioada de urmărire). Tariful orar al psihoterapeuţilor a fost stabilit la 6,59 USD, utilizând un principiu de echivalare bazat pe puterea de cumpărare ca modalitate de transformare a costurilor din lei într-o valută cu circulaţie internaţională. Nu sunt prevăzute costuri pentru servicii de psihiatrie decât în grupul celor care primesc fluoxetină. Costul fluoxetinei s-a stabilit pe baza preţului pieţei, luând în calcul preţul cel mai mic obţinut în baza a trei oferte obţinute de la distribuitori ai acestui produs. De asemenea, sunt prevăzute costuri specifice pentru cei care au realizat evaluări psihologice periodice necesare pentru a urmări evoluţia fiecărei persoane incluse în studiu. În anumite faze ale evaluării au participat psihologi clinicieni, iar

în altele evaluările au fost realizate de asistenţi de cercetare. Tarifele orare calculate se referă la venituri brute incluzând cheltuielile angajatorului şi reprezintă fracţiuni din venitul lunar mediu calculat la 168 de ore/lună. Costurile de regie se stabilesc procentual din totalul cheltuielilor directe necesare oferirii serviciului respectiv, fiind utilizate frecvent valori cuprinse între 25% şi 50% din totalul cheltuielilor directe. În cazul respectiv, s-a utilizat valoarea de 50%.

În privinţa costurilor de oportunitate s-au avut în vedere doar două cheltuieli: cele legate de timpul pacientului (clientului) şi de transport. În cazul timpului pacientului, pe lângă timpul efectiv destinat psihoterapiei, respectiv consilierii privind tratamentul medicamentos, se adaugă timpul necesar pentru deplasarea înspre şi de la locaţia unde este livrat serviciul primit[1]. În cazul deplasărilor, s-au normat 21, cu una în plus faţă de numărul de şedinţe de consiliere/psihoterapie urmate, fiind normată şedinţa iniţială de evaluare a situaţiei în pretest.

În privinţa costurilor implicate, tarifele specificate au fost cele utilizate. Acolo unde nu sunt disponibile asemenea date, pot fi estimate preţurile per unitate de măsură, luând în calcul preţul pieţei (de exemplu, cel mai mic preţ în urma unei cereri a trei oferte). Pentru transformarea lor în unităţi monetare cu circulaţie internaţională (de exemplu, euro sau dolari americani) se poate utiliza fie cursul de schimb mediu din perioada implementării unui proiect, fie paritatea bazată pe puterea de cumpărare (*purchasing power parity*).

Caseta 6.1. *Un model de chestionar de colectare a datelor privind costurile implicate*

Partea I. Chestionar de monitorizare a resursei de timp (completat de terapeut, cu ajutorul informaţiilor oferite de părintele însoţitor)

Numele copilului _____________

Nr. şedinţei	A. Timpul alocat de către specialist			C. Timpul persoanei care însoţeşte copilul			C. Alte costuri de oportunitate	
	A1 Timpul de pregătire (min.)	A2 Şedinţa de terapie (min.)	A3 Managementul cazului (min.)	C1 Teme pentru acasă (min.)	C2 Şedinţa de terapie (min.)	C3 Durata de deplasare (min.)	D1 Costul deplasării dus-întors	D2 Alte costuri de îngrijire (de exemplu, bonă)
Şedinţa 1								

Şedinţa 2
…
Şedinţa n
TOTAL

Partea a II-a. Chestionar de estimare a altor costuri de oportunitate (completat de părintele însoţitor)

1. Sunteţi angajat în prezent?
   a) Da, cu normă întreagă.
   b) Da, cu normă parţială.
   c) Nu, momentan nu lucrez.
   d) Altă situaţie, specificaţi care? ___________________________
2. Estimaţi cu referire la ultimele două luni, cât de mult timp v-a luat din timpul Dvs. de muncă sau din timpul Dvs. liber să rezolvaţi probleme ale copilului Dvs. cum ar fi:

I. Vizite la şcoală pentru a discuta cu profesorii/direcţiunea probleme ale copilului Dvs.
   a) deloc;
   b) 1-2 ore;
   c) 3-4 ore ;
   d) Mai mult de 4 ore. Câte? _____________

II. Vizite la medic din cauza unor accidentări suferite de copilul Dvs.
   a) deloc;
   b) 1-2 ore;
   c) 3-4 ore ;
   d) Mai mult de 4 ore. Câte? _____________

3. Ce cheltuieli aţi avut în ultimele două luni cu copilul Dvs. (exprimaţi suma în lei sau încercuiţi fără cheltuială, după caz):
   e) Cheltuieli cu asistenţă medicală şi medicamente pentru tratarea ADHD sau pentru tratarea unor răni/traumatisme determinate de neatenţia/hiperactivitatea copilului Dvs. în valoare de ___________/fără cheltuieli;
   f) Angajarea unui tutore privat pentru a face faţă lecţiilor de la şcoală, cheltuieli în valoare de ____________/fără cheltuieli;
   g) Cheltuieli cu acoperirea unor daune materiale produse de copilul Dvs. în valoare de _____________/fără cheltuieli.

Din cele prezentate referitoare la costuri rezultă câteva observaţii şi recomandări generale: (a) sub umbrela unui proces aparent riguros şi obiectiv, precum este analiza cost-eficienţă, se găsesc o serie de decizii subiective, unele cu impact semnificativ asupra rezultatelor obţinute (de exemplu, se ajunge la anumite costuri dacă se apelează la tariful orar calculat ca fracţiune din cele 168 de ore

corespunzătoare unui salariu lunar şi alte costuri dacă tariful orar are în vedere valorile practicate în cabinete private de către psihoterapeuţii cu un prestigiu recunoscut); (b) ori de câte ori este posibil, se iau în calcul costurile reale ce au fost implicate, iar acolo unde acest lucru nu este posibil, estimările trebuie să fie realiste, bazate preponderent pe un studiu al pieţei; (c) pentru a reflecta impactul unor asemenea decizii se recomandă ca rezultatul final obţinut să fie supus unei analize de sensibilitate, ce constă în analiza robusteţii rezultatului obţinut în contextul luării în considerare a mai multor scenarii de costuri (de exemplu, care ar fi fost rezultatul dacă tariful orar mediu al unei şedinţe de psihoterapie ar fi fost de 20 de USD); (d) pentru o estimare adecvată a costurilor şi resurselor, dincolo de problematica preţului, este esenţial să se identifice clar toate secvenţele şi resursele implicate (consumate).

6.2.2. Estimarea eficienţei

Estimarea eficienţei se realizează pornind de la rezultatele (efectele) pe care serviciul implementat le-a avut în urma derulării studiului respectiv. Dacă studiul are în vedere eficienţa unor intervenţii pentru reducerea greutăţii corporale, atunci eficienţa vizează kilogramele reduse. Dacă intervenţia îşi propune să reducă nivelul de epuizare emoţională, eficienţa vizează intensitatea scăderii punctajului dinspre un nivel care indică un nivel intens de *burnout* spre un nivel scăzut sau foarte scăzut al nivelului de *burnout*. Dacă programul vizează prevenirea fumatului în rândul adolescenţilor, efectul urmărit ar putea fi numărul de adolescenţi avertizaţi (informaţi). În schimb, dacă programul vizează renunţarea la fumat, efectul ar putea fi numărul de adolescenţi fumători care au renunţat la acest viciu. Prin urmare, în alegerea indicatorilor utilizaţi pentru a măsura eficienţa unei intervenţii, un loc central îl revine obiectivului urmărit prin intervenţia respectivă.

Prima măsură utilă cu referire la analiza comparativă a eficienţei unor intervenţii psihologice este oferită de rezultatele statistice clasice. Dacă sunt identificate diferenţe semnificative statistic între intervenţii, putem concluziona că nu au acelaşi nivel de eficienţă. De asemenea, dacă diferenţele dintre grupele comparate nu sunt semnificative statistic, iar distanţele dintre mediile grupelor sunt

mici, se poate spune că serviciile comparate au acelaşi nivel de eficienţă (vezi referirile din capitolul 4, tabelul 4.4, privind demonstrarea non-inferiorităţii sau a echivalenţei eficienţei diferitelor forme de intervenţie).

Dincolo de această analiză iniţială, se apelează frecvent la transformarea scorurilor brute într-o altă formă cu sens practic inclusiv pentru nespecialişti. De exemplu, în cazul depresiei, pentru oamenii de decizie nespecialişti în problematica sănătăţii mentale, un scor de 15 la o probă clasică de depresie nu are multă semnificaţie. Ei sunt interesaţi, mai degrabă, de exprimarea unor rezultate direct legate de scopul urmărit al programului de intervenţie. Pentru ei nu contează punctajul brut obţinut, ci dacă în urma intervenţiei individul a fost recuperat sau nu, dacă mai suferă sau nu de depresie. De aceea, ori de câte ori efectul urmărit nu reprezintă un rezultat important pentru comunitate (cum ar fi: recidivează sau nu; este sănătos sau nu; pleacă sau nu din companie în cel mult un an de la angajare etc.), se apelează la acest pas de transformare a rezultatelor brute în expresii cu sens recunoscut la nivel social.

De exemplu, în studiul amintit (Sava *et al.*, 2009), eficienţa a fost măsurată prin scorurile la scala de depresie Beck, însă acestea au fost transformate în rezultate cu sens, cum ar fi numărul de zile fără simptomatologie depresivă (*depression free days)*. La un astfel de *output* se poate ajunge apelând la următorii paşi standard: (a) raportarea scorurilor brute la etalonul probei pentru a vedea care dintre persoane mai suferă de simptomatologia specifică depresiei majore; (b) transformarea pe baza unui algoritm precum interpolarea liniară a intensităţii simptomelor în număr de zile cu manifestări depresive.

În cazul respectiv, consultând etalonul probei Beck (David, 2006), se observă că scoruri la BDI ≤ 7 înseamnă însănătoşire totală (absenţa simptomatologiei depresive), în timp ce scorurile BDI ≥ 29 înseamnă o simptomatologie depresivă severă. Valorile cuprinse între aceste limite indică prezenţa unei simptomatologii depresive dinspre depresie uşoară spre depresie moderată. Pornind de la aceste valori standard, s-a utilizat următorul algoritm de transformare în zile fără simptomatologie depresivă, luând ca reper o perioadă de 30 de zile. Celor care au obţinut un scor BDI ≤ 7 li s-a atribuit o pondere egală cu 1, considerându-se că şi-au păstrat această stare psihică în ultimele 30 de zile (ceea ce înseamnă automat 0

zile cu simptomatologie depresivă sau 30 de zile fără simptomatologie depresivă). Celor care au obţinut un scor BDI ≥ 29 li s-a atribuit o pondere egală cu 0. Utilizând aceeaşi asumpţie a menţinerii constante a stării mentale pentru ultimele 30 de zile, se ajunge, în cazul lor, la 0 zile fără simptomatologie depresivă sau la 30 de zile cu depresie. Pentru scorurile BDI cuprinse între 8 şi 28 se poate stabili numărul de zile fără simptomatologie depresivă cu ajutorul unei simple operaţii de interpolare liniară. De exemplu, în cazul unei persoane care avea un scor BDI de 10, ponderea acordată este de 0,86 în baza formulei folosite[2]. Pentru calcularea numărului de zile arondate fără simptomatologie depresivă, se înmulţeşte această pondere cu perioada de timp luată în calcul (în acest caz 30 de zile), obţinându-se 25,8 zile fără simptomatologie depresivă raportate din cele 30 de zile anterioare momentului evaluării, acest lucru fiind atribuit unei persoane cu un scor BDI = 10[3].

Desigur, un asemenea algoritm de lucru este necesar atunci când variabila dependentă principală este măsurată numeric şi este exprimată într-o formă mai puţin înţeleasă de către publicul larg şi de către decidenţii programelor educaţionale sau de sănătate vizate (vezi discuţia arondată tabelului 4.6 din capitolul 4, cu privire la distincţia dintre VD surogat şi cele reale). Ipotetic, există alte trei posibile scenarii: (a) variabile-surogat măsurate dihotomic (intenţionează sau nu să părăsească compania); (b) variabile reale măsurate dihotomic (fluctuaţia reală observată operaţionalizată dihotomic: a plecat sau nu din companie); (c) variabile reale măsurate numeric (volumul lunar al vânzărilor realizate). În cazurile cu VD brute exprimate într-o manieră dihotomică (a şi b), nu se pretează utilizarea interpolării liniare, iar în cazurile cu VD reale (b şi c) nu este necesar pasul de transformare a rezultatului brut în alte expresii, deoarece la nivel social asemenea rezultate (fluctuaţia de personal, volumul lunar al vânzărilor etc.) au sens în mod direct.

Deşi nu există un singur indicator de exprimare a eficienţei unui serviciu psihologic, cel puţin în zona intervenţiilor psihologice cu aplicabilitate în domeniul clinic şi al sănătăţii se utilizează frecvent un indicator temporal (cantitativ) al problematicii investigate, cum ar fi zilele fără simptomatologie depresivă, zilele fără simptomatologie ADHD etc. Există cel puţin două avantaje majore ale utilizării unui asemenea indicator (Vannoy *et al.*, 2010): (a) facilitează realizarea altor tipuri

de analize economice cum ar fi analiza cost-utilitate (deoarece indicatorul cantitativ de timp cu/fără simptomatologie este un pilon important în calcularea indicatorilor standard tipici în asemenea analize, cum este QALY (*quality adjusted life years*); (b) facilitează scorurile agregate ale eficienţei unei intervenţii atunci când VD este măsurat la diferite intervale de timp (de exemplu, înainte, imediat după intervenţia cu o durată de trei luni, dar şi la şase, respectiv 12 luni de la finalizarea intervenţiei, ca perioadă de urmărire a stabilităţii efectelor). Eficienţa unui program este egală cu media ponderată a eficienţei (de exemplu, numărul ZFS calculat pentru fiecare individ în parte care a urmat acel program), aşa cum se observă în caseta 6.2.

Caseta 6.2. *Exemple concrete de calcul a zilelor fără simptomatologie (ZFS)*

În cazul în care sunt patru evaluări: înainte de implementarea unei intervenţii cu durata de 3 luni (T0), imediat după finalizarea intervenţiei (T1 = T0 + 3 luni), în perioada de urmărire la 3 luni de la finalizarea intervenţiei (T2 = T1 + 3 luni) şi în perioada de urmărire la 9 luni de la finalizarea intervenţiei (T3 = T2 + 6 luni), intensitatea simptomatologiei conduce la anumite ponderi utilizate în calculul pentru stabilirea numărului de zile fără simptomatologie (ZFS). Aşa cum am amintit, aceste ponderi sunt cuprinse între 0 (dacă problema a fost rezolvată) şi 1 (dacă problema este una severă). În toate perioadele urmărite se poate calcula o pondere medie în baza evaluărilor existente în acea perioadă. De exemplu, pentru a afla numărul ZFS dintre T0 şi T1, se calculează media celor două ponderi obţinute cu ocazia primelor două evaluări şi se înmulţeşte acest rezultat cu perioada avută în vedere (în acest caz, 90 de zile). De exemplu, dacă un participant într-un studiu primeşte ponderea 0 (simptomatologie severă) la momentul T0, respectiv ponderea 1 (recuperare completă) la momentul T1, se obţine o medie de 0,50 care înmulţită cu perioada de 90 de zile înseamnă 45 de ZFS în cazul său. Dacă există mai mult de două evaluări în intervalul avut în vedere (de exemplu, se doreşte stabilirea numărului de ZFS dintr-un an de zile, cu cele patru momente de evaluare), formula devine: ZFS = 365 zile × [(0,125 × ZFS T0) + (0,250 × ZFS T1) + (0,375 × ZFS T2) + (0,250 × ZFS T3)]. Spre exemplificare, ponderea de 0,125 pentru T0 este determinată prin faptul că rezultatele acestei evaluări reprezintă jumătate din evaluarea unei singure perioade cu durata de 3 luni (durată ce reprezintă 25% dintr-un an, între T0 şi T1). Ponderea de 0,375 acordată lui T2 rezultă din însumarea lui 0,125 (jumătate din perioada de 3 luni cu pondere 0,250 dintre T1 şi T2) şi 0,250 (jumătate din perioada de 6 luni cu pondere de 0,500 dintr-un an, dintre T2 şi T3). Însumate, toate aceste ponderi trebuie să fie egale cu 1,00, valorile lor fiind stabilite în funcţie de numărul de evaluări existente şi de distanţa temporală dintre evaluări raportată la durata totală a perioadei avute în vedere. De exemplu, dacă ar fi avut loc doar trei evaluări, T0, T1 şi T2 cu o durată totală a perioadei analizate de 6 luni, formula devine ZFS: 182 de zile × [(0,25 × ZFS T0) + (0,50 × ZFS T1) + (0,25 × ZFS T2)].

Există o serie de neajunsuri ale acestui model, cel mai important fiind vulnerabilitatea asumpţiei că modificarea simptomatologiei are o formă liniară. De exemplu, între T0 şi T1 se consideră că are loc o îmbunătăţire continuă, deşi în realitate ar putea fi vorba de o evoluţie neliniară, cu scenarii multiple: (a) scenariul ameliorării produse abia spre finalul perioadei, ceea ce ar presupune o

supraestimare a ZFS; (b) scenariul în care evoluţia este una curbilinie, după începerea terapiei fiind o perioadă de intensificare a simptomelor, iar abia apoi apare o ameliorare (acelaşi risc de supraestimare a ZFS); (c) scenariul evoluţiei ciclice, cu o ameliorare a simptomatologiei între măsurători rămasă nesurprinsă, însoţită de o recădere în apropierea T1 (efect de supraestimare a ZFS).

Aceste critici sunt cu atât mai îndreptăţite cu cât există o distanţă temporală mai mare între momentele evaluării, îndeosebi între T0 şi T1, în care se oferă intervenţia[4]. O soluţie parţial compensatorie constă în restrângerea perioadei de transformare a ZFS cu scopul de a minimiza distorsiunile. De exemplu, Sava *et al.* (2009) ia în calcul un interval limitat de 30 de zile anterioare momentului evaluării pentru calculul ZFS (ZFS pentru 30 de zile anterioare lui T0; ZFS pentru 30 de zile anterioare lui T1 – măsurat la 3 luni distanţă faţă de T0, ZFS pentru 30 de zile anterioare lui T2, măsurătoare făcută la 6 luni distanţă faţă de T1). În acest caz, se compară valorile ZFS lunare înainte (T0), imediat după intervenţie (T1) şi la şase luni de la finalizarea intervenţiei (T2), fără a se calcula un ZFS total pentru toată perioada.

În sfârşit, modalitatea de calcul a eficienţei unei intervenţii se complică dacă nu există clar o VD principală. De exemplu, din perspectiva unui program finanţat de stat pentru recuperarea după un traumatism sever al mâinii şi al antebraţului, din perspectivă psihosocială, pe lângă reducerea distresului emoţional (depresie şi/sau anxietate), poate fi foarte importantă o reintegrare profesională cât mai rapidă. În asemenea cazuri, deşi avem un singur cost, pot exista multiple rezultate pozitive. În asemenea situaţii, se pot calcula separat analize de tip cost-eficienţă pentru fiecare beneficiu major sau se poate calcula un indicator agregat al eficienţei în funcţie de ponderea acordată fiecărui efect urmărit. În exemplul oferit, dacă programul psihoterapeutic urmărit are ca obiectiv reducerea distresului emoţional, se pot acorda următoarele ponderi: 0,40 pentru depresie, 0,40 pentru anxietate şi 0,20 pentru reintegrare profesională, după care se poate obţine şi o valoare globală a ZFS, utilizând ponderile respective (vezi Yates, 1999).

6.2.3. Analiza cost-eficienţă

După estimarea costurilor implicate şi a eficienţei serviciilor psihologice se pot asambla cele două componente pentru a realiza o analiză de tip cost-eficienţă. Unitatea de bază este raportul cost-eficienţă, în cadrul căreia cu cât costurile per unitate de eficienţă sunt mai reduse, cu atât intervenţia analizată are un raport cost-eficienţă mai bun. În principiu, dacă intervenţia A are un raport cost-eficienţă pentru o zi fără simptomatologie (de 20 USD/ZFS) sau pentru un an fără simptomatologie (AFS) (de 7.300 USD/AFS), se poate spune că ea este superioară intervenţiei B sub raport cost-eficienţă, dacă raportul cost-eficienţă al acesteia este de 50 USD/ZFS, respectiv de 18.250 USD/AFS.

În general, se poate vorbi despre trei tipuri de analiză cost-eficienţă: (a) analiza mediei raportului cost-eficienţă (ACER – *average cost-effectiveness ratio*) care se

aplică atunci când se compară eficienţa unei intervenţii cu situaţia absenţei intervenţiei respective (comparaţia dintre pretest şi posttest, respectiv comparaţia dintre grupul care primeşte intervenţia şi un grup de control pasiv); (b) analiza raportului marginal cost-eficienţă (MCER – *marginal cost-effectiveness ratio*), care vizează acelaşi lucru în condiţiile varierii costurilor ca urmare a extinderii sau reducerii unui program; (c) şi analiza raportului incremental cost-eficienţă (ICER – *incremental cost-effectiveness ratio*), aplicabilă atunci când sunt comparate două sau mai multe intervenţii active între ele (Denil *et al.*, s.a.).

În cazul ACER, de exemplu, se vorbeşte despre un raport cost-eficienţă pentru fiecare ZFS obţinută în plus ca urmare a intervenţiei. Bunăoară, în cazul unei intervenţii care a costat 500 USD/client şi care a condus la un câştig de 20 de ZFS, de la 3 ZFS înainte de intervenţie la 23 ZFS după intervenţie), raportul mediu cost-eficienţă ar fi de 25 USD/ZFS câştigată, respectiv de 9.125 USD/AFS câştigat. Aceste rezultate prezentate constituie valori descriptive medii obţinute. Dacă asemenea raporturi sunt calculate la nivel individual, pentru fiecare persoană inclusă în cercetare, luând în calcul costurile şi eficienţa observată pentru fiecare persoană evaluată, se poate apela la tehnici statistice inferenţiale clasice, pentru a testa dacă există sau nu diferenţe semnificative statistic în privinţa raportului cost-eficienţă între anumite forme de intervenţie[5]. De exemplu, în studiul realizat de Sava *et al.* (2009), utilizând o asemenea strategie, s-a constatat că apelul la psihoterapie, fie ea sub forma REBT (M = 23,77 USD/ZFS câştigată), sau CT (M = 26,22 USD/ZFS câştigată), conduce la un raport cost-eficienţă semnificativ mai scăzut decât în cazul apelului la medicaţie (M = 34,93 USD/ZFS câştigată). Analizele *post hoc* nu au găsit diferenţe semnificative între cele două forme de psihoterapie.

În anumite situaţii se poate apela la analize de tip MCER, pentru a urmări impactul asupra raportului cost-eficienţă, dacă anumite componente implicate în calculul costurilor se modifică semnificativ. De exemplu, în cazul tratamentului cu fluoxetină este posibil să scadă costurile pe persoană cu medicaţia în cazul în care în loc de 50 de clienţi sunt tratate 500 de persoane, datorită posibilităţii de a negocia un preţ mai bun pentru medicamente în cazul unui volum solicitat de 10 ori mai mare decât în cel utilizat pentru a calcula costul medicaţiei. În schimb, dacă preţul

psihoterapiei, aşa cum a fost el calculat ca fracţiune din venitul salarial lunar nu s-ar reduce, ar putea afecta rezultatul raportului cost-eficienţă menţionat anterior. Analizele MCER au în vedere asemenea situaţii. Totuşi, acest gen de analiză este rar întâlnit, fiind preferată includerea variaţiilor posibile de costuri şi a implicaţiilor acestora în analize de sensibilitate, descrise în secţiunea 6.2.5.

O strategie frecvent întâlnită în analiza raportului cost-eficienţă, valabilă pentru programe concurente, este aceea de a calcula un raport incremental cost-eficienţă (ICER) (Phillips, 2009). Formula în acest caz este raportul dintre diferenţa de costuri dintre două intervenţii (C2 – C1) şi diferenţa de eficienţă dintre aceleaşi două intervenţii (E2 – E1). În cazul în care se compară mai mult de două intervenţii, înainte de utilizarea formulei respective se ierarhizează intervenţiile în ordine crescătoare a eficienţei lor (vezi exemplul ipotetic al tabelelor 6.3 şi 6.4).

Revenind la exemplul nostru privind studiul asupra depresiei majore, din tabelul 6.3 se observă că apelul la terapie cognitivă (CT) conduce la un ICER cu semn negativ. Acest lucru înseamnă că adoptarea CT faţă de intervenţia precedentă (fluoxetina), conduce la obţinerea unui efect incremental (mai multe ZFS) cu bani mai puţini. Prin urmare, fluoxetina este eliminată din analiză, procesul stabilirii raporturilor ICER reluându-se într-o manieră iterativă cu celelalte intervenţii rămase (vezi tabelul 6.4).

Tabelul 6.3. *Raporturile ICER (raporturile incrementale cost-eficienţă). Pasul 1*

Intervenţie	Cost	Efect (eficienţă)	Δ Costuri	Δ Eficienţă	ICER (ΔC/ΔE)
Imipramină	495 USD/om	15 ZFS/câştigate lunar	495 USD	15	33
Fluoxetină	665 USD/om	18 ZFS/câştigate lunar	170 USD	3	21,66
CT	505 USD/om	19 ZFS/câştigate lunar	–160 USD	1	–160
REBT	532 USD/om	20 ZFS/câştigate lunar	27 USD	1	27

Tabelul 6.4. *Raporturile ICER (raporturile incrementale cost-eficienţă). Pasul 2*

Intervenţie	Cost	Efect (eficienţă)	Δ Costuri	Δ Eficienţă	ICER (ΔC/ΔE)
Imipramină	495 USD/om	15 ZFS/câştigate lunar	495 USD	15	33,0
CT	505 USD/om	19 ZFS/câştigate lunar	10 USD	4	2,5
REBT	532 USD/om	20 ZFS/câştigate lunar	27 USD	1	27,0

În urma analizei din tabelul 6.4 va fi eliminată şi alternativa imipraminei, deoarece aceasta este mai puţin eficientă decât CT (15 *vs* 19 ZFS) şi costă mai mult pentru a produce un ZFS suplimentar (33 *vs* 2,5)[6]. În sfârşit, în tabelul 6.5 sunt analizate comparativ cele două programe rămase în cursă, CT şi REBT.

Tabelul 6.5. *Raporturile ICER (raporturile incrementale cost-eficienţă). Pasul 3*

Intervenţie	Cost	Efect (Eficienţă)	Δ Costuri	Δ Eficienţă	ICER (ΔC/ΔE)
CT	505 USD/om	19 ZFS/câştigate lunar	505 USD	19	26,57
REBT	532 USD/om	20 ZFS/câştigate lunar	27 USD	1	27,00

Din tabelul 6.5 se observă că ambele programe au raporturi cost-eficienţă similare. Astfel, cu toate că REBT pare a avea o eficienţă suplimentară (20 ZFS *vs* 19 ZFS), costul per unitate suplimentară de eficienţă este de asemenea uşor mai ridicat (27 *vs* 26,57). În asemenea condiţii, se poate concluziona că cele două programe au niveluri similare ca raport cost-eficienţă, recomandarea de a utiliza una sau alta dintre formele de intervenţie fiind dependentă de bugetul avut la dispoziţie. Dacă bugetul este suficient de generos pentru a acoperi cheltuielile tuturor cazurilor de depresie majoră, recomandarea ar putea merge către REBT, datorită plusului de eficienţă (dacă există diferenţe semnificative statistic în privinţa eficienţei celor două forme de intervenţie). Dacă asemenea diferenţe nu există sau dacă bugetul nu permite acoperirea tuturor cazurilor de depresie majoră, se alege intervenţia cu costuri mai reduse (dacă există o diferenţă semnificativă statistic sub raportul costurilor implicate). În cazul în care ambele programe au niveluri similare de costuri, respectiv de eficienţă, se poate apela la oricare dintre cele două forme de intervenţie.

6.2.4. Ilustrarea unei analize cost-eficienţă în domeniul organizaţional

Exemplul ales este al determinării celei mai bune modalităţi de instruire a noilor angajaţi dintr-o companie multinaţională aflată în extindere, cu scopul de a obţine o integrare cât mai rapidă a lor la locul de muncă (de obicei, noii angajaţi au nevoie

de o perioadă de acomodare până să ajungă la un nivel maxim de productivitate – operaţionalizată prin atingerea sau chiar depăşirea normei zilnice). Deşi acest gen de studii se pretează şi la analize de tip cost-beneficiu, ne vom rezuma în această secţiune la o analiză de tip cost-eficienţă. Criteriul ales pentru eficienţă este numărul de zile în care norma a fost atinsă pe o perioadă de urmărire de trei luni de zile, cu patru perioade de evaluare: în a prima zi efectivă de muncă, în a 15-a zi, în a 30-a zi şi în ultima zi supusă monitorizării, a 90-a. Cele trei variante supuse analizei de cost-eficienţă[7] au fost următoarele:

(a) 3 zile de training de iniţiere, fiecare cu durata de 8 ore pe zi, susţinute de personal propriu (serviciu intern); trainingul s-a derulat imediat după angajare, fiind necesare 10 asemenea şedinţe de instruire (grupuri de câte 10) pentru a-i integra pe cei 100 de noi angajaţi;

(b) 2 zile de training de iniţiere, fiecare cu durata de 8 ore pe zi, susţinute de traineri externi (serviciu externalizat), imediat după angajare, fiind estimate tot 10 asemenea şedinţe de instruire (grupuri formate din 10 persoane, totalizând 100 de noi angajaţi);

(c) on site training (desemnarea unor colegi experimentaţi care să-i instruiască individual pe noii angajaţi la locul lor de muncă, timp de o oră/zi lucrătoare în prima lună, respectiv timp de 15 minute/zi lucrătoare în lunile 2-3; sunt estimate 21 de zile lucrătoare şi 100 de noi angajaţi).

6.2.4.1. Determinarea costurilor

Dacă ne referim, ilustrativ, la programul de instruire cu traineri interni, am putea include în cadrul costurilor directe de personal (1): (a) veniturile angajaţilor desemnaţi să ţină acest training, fracţionat ca timp de lucru pe care îl petrec pentru a derula sesiunile respective, dar şi pentru a pregăti materialele de instruire[8]; (b) veniturile nou-angajaţilor pe durata implicării lor în training, fiind vorba de timpi neproductivi); (c) veniturile echipei de suport, fracţionat în funcţie de durata de timp alocată de aceştia pentru pregătirea celor necesare trainingului (personal administrativ etc.); (d) costul eventual al unor servicii de consultanţă, dacă este cazul. În cazul veniturilor salariale se iau în considerare veniturile brute la care se

adaugă cheltuielile angajatorului. Tot în categoria costurilor directe avem cheltuielile logistice (2) (de exemplu, închirierea aparaturii audio-video, *flip-chart*, hârtie, costuri pentru multiplicare, închirierea spaţiului etc., fiecare dintre acestea fiind determinate în mod realist, ţinând cont de condiţiile existente în firma care implementează sesiunile de training). A treia categorie importantă de costuri directe posibile sunt costurile referitoare la mobilităţi (3). Aici ar intra costurile cauzate de eventuala decizie de a ţine aceste întâlniri în alte locaţii (cazare, masă, transportul persoanelor şi echipamentelor, după caz). Pe baza acestor cheltuieli directe, în funcţie şi de elementele incluse sau excluse, se poate stabili şi o anumită valoare a costurilor de regie, stabilite ca procent din cheltuielile directe amintite mai sus, valoarea acesteia fiind adesea cuprinsă între 25% şi 50% (cuprinzând, după caz, costurile cu utilităţile, personalul administrativ, echipamentele puse la dispoziţie etc.). În general, valorile cheltuielilor de regie sunt mai mici dacă unii dintre itemii menţionaţi anterior au fost incluşi printre cheltuielile directe. La toate aceste cheltuieli se adaugă şi cheltuielile de oportunitate (cum ar fi pierderile în productivitate cauzate de implicarea celor instruiţi, dar şi a celor care instruiesc, într-o activitate care nu este direct productivă).

Tabelul 6.6. *Estimarea costurilor implicate în cele trei programe de instruire (date ipotetice)*

			Unităţi medii consumate per client			Totalul costului per client		
Tipul de resursă utilizat	Unitate de măsură	Preţ unitar $	Training externalizat	Training intern	*On site training*	Training externalizat	Training intern	*On site training*
Manoperă/timp (tarif orar)								
Angajat trainer	1 h/persoană *	0,60 $	0 h	26 h **	0 h	0 $	15,60 $	0 $
Trainer extern	1 h/persoană *	8,00 $	16 h	0 h	0 h	108,00 $	0 $	0 $
*S*upervizor *on site*	1h	5,00 $	0 h	0 h	25,25 h	0 $	0 $	126,25 $
Angajat instruit	1h	4,50 $	16 h	24 h	25,25 h	72,00 $	108,00 $	113,62 $
Închirierea	1	20,00	1 sală	1 sală	0 săli	20,00 $	20,00 $	0 $

spaţiului cu dotat cu aparatura necesară	sală/training/persoană *	$						
Cheltuieli cu pauza de cafea	1 pauză/persoană	2,00 $	4 pauze	6 pauze	0 pauze	8,00 $	12,00 $	0 $
Consumabile – dosar de training	1 set/persoană	5,00 $	1 set	1 set	0 seturi	5,00 $	5,00 $	0 $
Cheltuieli cu masa de prânz	1 masă/persoană	5,00 $	2 mese	3 mese	0 mese	10,00 $	15,00 $	0 $
Totalul cheltuielilor directe (CD)						*223,00* $	*175,60* $	*239,87* $
Regie (25% din CD)						*55,75* $	*43,90* $	*59,96* $
Totalul costului serviciului						***278,75*** $	***219,50*** $	***299,83*** $
Costul pierderii de productivitate***	1 h/persoană	4,50 $	16 h	24 h	25,25 h	72,00 $	108,00 $	113,62 $
Cheltuieli deductibile – manoperă	factură/persoană	17,28 $	0 nefacturat	1 nefacturat	1 nefacturat	0 $	17,28 $	17,28 $
Cheltuieli deductibile – logistică	factură/persoană	6,88 $	0 nefacturat	0 nefacturat	1 nefacturat	0 $	0 $	6,88 $
Totalul costului de oportunitate						***72,00*** $	***125,28*** $	***137,78*** $
COSTUL TOTAL PE ANGAJAT	(Costul serviciului + Costul de oportunitate)/CLIENT					**350,75** $	**344,78** $	**437,61** $
COSTUL INTERVENŢIEI	= Nr. oamenilor instruiţi × Costul total pe angajat instruit					**35.075** $	**34.478** $	**43.761** $

* Am calculat cât revine costurilor cu închirierea sălii de training dotate corespunzător, raportată la o singură persoană (fiind instruite câte 10 persoane per training), pentru a permite raportarea cheltuielilor per individ. Acelaşi lucru a fost făcut şi pentru traineri (în cadrul cheltuielilor cu manopera, cu excepţia cazului de training *on site*, în care instruirea s-a făcut individual), precum şi pentru cheltuielile facturate care reduc valoarea impozitului pe profit (provenite din facturarea cheltuielilor cu trainerul extern şi a celor logistice, calculate la 16%).

** În cazul instructorilor angajaţi interni, pe lângă cele 24 de ore cumulate de training realizate efectiv în cele 3 zile de instruire au fost adăugate alte 2 ore/persoană ca urmare a unei estimări a 200 ore de muncă pentru conceperea şi pregătirea trainingului derulat pe cele 100 de persoane.

*** În estimarea pierderilor legate de productivitate s-a ţinut cont de cheltuielile în regim plata cu ora, activităţile derulate în perioada de instruire fiind delegate către alţi colegi. Nu am luat în calcul, deşi ar fi fost cazul, diferenţele cauzate de eficienţa diferită a celor trei programe de instruire.

6.2.4.2. Determinarea efectelor

Obiectivul principal al programelor de instruire era de a reduce perioada de acomodare a noilor angajaţi astfel încât ei să atingă performanţele cerute în cadrul companiei (norma stabilită). Prin urmare, firesc ar fi ca principalul efect urmărit să fie reprezentat de numărul de zile, din cele 90 cuprinse în perioada de monitorizare, în care noii angajaţi reuşesc să-şi atingă norma. Rezultatele medii pentru cele trei programe au fost: 54 de zile pentru cei care au fost beneficiarii instruirii furnizate ca serviciu externalizat, 52 de zile pentru cei instruiţi de formatori interni, respectiv 63 de zile pentru cei care au urmat un programul de formare la locul de muncă (*on site*). Acestor date li se adaugă o statistică provenită de la alţi 25 de angajaţi noi, care din diverse motive nu au urmat niciuna dintre formele de instruire. Aceştia au atins norma de producţie stabilită în 40 de zile din cele 90 de zile stabilite, fără a implica vreun cost specific de instruire[9].

6.2.4.3. Analiza cost-eficienţă

Sintetizând datele prezentate anterior în tabelul 6.7, pot rezulta mai multe tipuri de analiză cost-eficienţă. Una dintre cele mai frecvente analize este cea a raportului incremental cost-eficienţă, utilă pentru a compara mai multe programe alternative între ele. Consultând datele din tabelul 6.7, se constată că varianta externalizării serviciului de instruire comparativ cu alternativa precedentă (trainingul intern) se dovedeşte a fi mai eficientă (există un efect incremental de 2 zile) în condiţiile unui cost/zi suplimentară câştigată mai redus (2,98 faţă de 6,74). După eliminarea alternativei de training intern (şi a absenţei oricărei intervenţii care nu este inclusă într-o analiză de tip ICER), în tabelul 6.8 putem urmări care dintre cele două alternative rămase în cursă are cel mai bun raport cost-eficienţă.

Tabelul 6.7. *Raporturi ICER (costul instruirii per zi cu normă îndeplinită)*

Intervenţie	*Cost*	*Efect (eficienţă)*	*Δ Costuri*	*Δ Eficienţă*	*ICER (ΔC/ΔE)*
Absentă	0 USD	40 de zile			
Training intern	344,78 USD	52 de zile	344,78 USD	12 de zile	28,73
Serviciu externalizat	350,75 USD	54 de zile	5,97 USD	2 zile	2,98
Instruire *on site*	437,61 USD	63 de zile	86,86 USD	9 zile	9,65

* Valorile ICER se calculează pentru a compara intervenţiile active. Absenţa programului de instruire se utilizează doar în calculul ACER (raportul clasic cost-eficienţă). Conform acestuia, surplusul de 12 zile cu normă îndeplinită prin instruire cu formatori interni costă 344,78 USD/om, adică 28,73 USD/zi în plus în care s-a îndeplinit norma. Surplusul de 14 zile, în cazul externalizării serviciului de instruire, are un cost de 350,75 USD/om instruit, adică 25,05 USD/zi în plus de îndeplinire a normei. În sfârşit, cele 23 de zile suplimentare cu normă îndeplinită obţinute prin training *on site* costă 437,61 USD/om, adică 19,02 USD/zi suplimentară cu norma îndeplinită. Din analiză pare a rezulta că instruirea *on site* are cel mai bun ACER. Pentru a fi siguri de acest lucru, se utilizează tehnici statistice comparative pentru a stabili dacă există diferenţe semnificative statistic sau se apelează la analize ICER.

Tabelul 6.8. *Raport ICER (costul instruirii per zi suplimentară cu normă îndeplinită)*

Intervenţie	*Cost*	*Efect (Eficienţă)*	*Δ Costuri*	*Δ Eficienţă*	*ICER (ΔC/ΔE)*
Serviciu externalizat	350,75 USD	54 de zile	350,75 USD	54 de zile	6,38
Instruire *on site*	437,61 USD	63 de zile	86,86 USD	9 zile	9,65

Conform acestui tabel, cele două intervenţii au un nivel similar al raportului cost-eficienţă. Deşi instruirea la locul de muncă are o eficienţă mai ridicată (există 9 zile suplimentare cu normă întreagă), costurile pe fiecare zi suplimentară cu normă îndeplinită sunt ceva mai mari (9,65 USD) faţă de situaţia apelului la un program de instruire iniţial oferit de traineri externi (6,38 USD). Reamintim că, în această analiză ICER, un program are un raport cost-eficienţă mai bun decât precedentul cu care este comparat în două situaţii: (a) dacă are o valoare ICER negativă (costă mai puţin şi are un efect mai bun) sau (b) dacă îndeplineşte o valoare ICER mai mică decât programul cu care este comparat şi are în acelaşi timp o eficienţă mai bună (cu alte cuvinte, dacă are un efect mai puternic şi un cost/efect incremental mai scăzut).

6.2.5. Analiza de sensibilitate

Orice raport cost-eficienţă poate varia mult în funcţie de costurile utilizate. Să ne gândim numai la un singur aspect: cum ar fi evoluat analizele dacă tariful pentru trainingul extern ar fi fost de 10 USD/oră în loc de 8 USD/oră? Costul total al intervenţiei ar fi crescut cu 40 de USD/om, ajungându-se la un cost de 390,75 USD/om, respectiv la un raport ICER de 22,98 în cadrul tabelului 6.7. În acest caz nu s-ar mai fi putut distinge între programele de instruire intern şi extern. În schimb, analiza ar fi arătat că instruirea la locul de muncă are un raport cost-eficienţă mai bun decât instruirea bazată pe serviciul externalizat, având o eficienţă mai bună şi un raport ICER mai redus ş.a.m.d.

În consecinţă, dată fiind importanţa identificării categoriilor de cost relevante şi a preţului corect pentru acestea, se recomandă prezentarea intervalelor de încredere asociate rezultatelor, cel puţin cu privire la raportul cost-eficienţă, fie el exprimat ACER sau ICER (Yates, 1999; Phillips, 2009).

Forma incipientă a analizei de sensibilitate presupune utilizarea intervalelor de încredere, posibil de calculat dacă datele sunt colectate individual cu privire la costurile intervenţiei şi/sau la efectele observate. În această direcţie vor fi favorizate programele cu efecte constante în dauna celor cu rezultate eterogene, deoarece vor conduce la intervale de încredere cu amplitudine mai mică. Aceste intervale de încredere pot fi stabilite şi prin apelul la tehnica *bootstrap*.

Formele intermediare ale analizei sensibilităţii urmăresc implicaţiile diferitelor asumpţii asupra rezultatelor evaluării. Aici intră efectul contextului (de exemplu, perioada de timp avută în vedere: se obţin aceleaşi rezultate dacă analiza se face pe 60 de zile în loc de 90 de zile?; numărul de persoane instruite: cum ar evolua costurile per participant dacă ar fi instruiţi doar 50 în loc de 100 de noi angajaţi?). De asemenea, aici intră analiza impactului diferitelor costuri asupra rezultatului analizei: „Cum ar evolua costurile celor două programe care necesită spaţiu special de instruire dacă acesta, împreună cu echipamentele necesare, sunt accesate gratuit de către companie?”, „Cum ar evolua rezultatele dacă instruirea la locul de muncă ar fi condensată în doar 45 de minute/zi în prima lună şi câte 10 minute/zi în celelalte două luni, adică ar dura cu şase ore mai puţin (19,25 faţă de 25,25 ore)?” ş.a.m.d. Aşa cum am precizat la începutul secţiunii 6.2.3, acest tip de analiză implică apelul la MCER (*marginal cost-effectiveness analysis*), centrată pe

optimizarea costurilor ca urmare a modificării (extinderii sau reducerii) unui serviciu analizat.

În sfârşit, în forma sa avansată (Phillips, 2009), analiza de sensibilitate implică o analiză de probabilităţi pentru a stabili curba de acceptabilitate a raportului cost-eficienţă. Valorile costurilor şi cele ale efectelor sunt utilizate pentru a genera eşantioane noi (de obicei 1.000 de noi eşantioane), obţinându-se pentru fiecare eşantion un raport cost-eficienţă. În funcţie de aceste valori se stabileşte o probabilitate de obţinere a unui anumit raport cost-eficienţă. De exemplu, fie o probabilitate de 0,25 ca raportul cost-eficienţă (ACER) să fie de cel mult 100 USD/zi câştigată cu normă întreagă, respectiv o probabilitate de 0,98 ca raportul să fie de cel mult 300 USD/zi câştigată cu normă îndeplinită (curba de acceptabilitate implică valorile CER pe abscisă şi probabilităţile pe ordonată). Interpretarea se face prin raportarea la anumite valori standard (*benchmarks*), considerate a fi acceptabile. De pildă, dacă conducerea companiei consideră că nivelul maxim al costului suportat pentru o zi cu normă întreagă este de 250 USD, iar probabilitatea de a avea un asemenea raport este de 0,90, există şanse de a accepta intervenţia respectivă, deoarece există de 9 ori mai multe şanse ca intervenţia să aibă un raport cost-eficienţă acceptabil decât unul inacceptabil. În schimb, dacă conducerea este dispusă să plătească maximum 100 USD/zi câştigată cu normă îndeplinită, programul nu va fi adoptat, fiind o probabilitate scăzută (0,25) ca raportul cost-eficienţă să fie în limitele acceptabile, de cel mult 100 USD (altfel spus, sunt de patru ori mai puţine şanse ca intervenţia să aibă un raport cost-eficienţă acceptabil decât unul inacceptabil). Mai multe detalii despre ce presupune analiza de sensibilitate sunt prezentate într-un limbaj accesibil nespecialiştilor în Taylor (2009).

6.3. Analiza cost-utilitate

Analiza cost-utilitate se apropie mult de analiza cost-eficienţă. Atât costurile, cât şi diferitele raporturi de tip cost-utilitate (de exemplu, ICER) sunt similare cu cele întâlnite în cazul analizei cost-eficienţă. Ceea ce diferă este efectul măsurat. Într-o

analiză cost-utilitate (CUA – *cost-utility analysis*) efectul este standardizat, ceea ce permite comparaţii între intervenţii, chiar dacă acestea se referă la probleme diferite (Yates, 1999; McCabe, 2009; Whitehead şi Shehzad, 2010). De exemplu, prin analize de tip cost-utilitate se poate urmări care dintre următoarele intervenţii are un raport cost-utilitate mai bun: (a) terapia cognitiv-comportamentală pentru combaterea depresiei majore; (b) utilizarea hipnoterapiei pentru reducerea durerii postoperatorii sau (c) coaching-ul directiv pentru reducerea nivelului de *burnout* la angajaţi.

Acest avantaj al CUA este important atunci când există constrângeri de natură bugetară care nu permit finanţarea simultană a tuturor acestor programe ce urmăresc probleme diferite. Pentru a realiza acest lucru, analiza cost-utilitate recurge la conceptul de utilitate, existând mai multe căi de obţinere a unei utilităţi asociată fiecărei stări de sănătate. Un asemenea artificiu are în vedere transformarea efectului măsurat în utilitate. Conform analizei cost-utilitate, nu se poate pune un semn de egalitate între două intervenţii care au costuri egale şi care obţin aceleaşi rezultate în termeni de ZFS (zile fără simptomatologie), dacă într-un caz ZFS se referă la recuperarea din depresie, iar în celălalt la recuperarea din arahnofobie (teamă de păianjeni). Inegalitatea în efecte se produce la nivel valoric, deoarece revenirea dintr-o stare depresivă are mai multe implicaţii asupra funcţionării eficiente la nivel psihic, social şi profesional decât revenirea din arahnofobie. Prin urmare, analiza cost-utilitate se raportează la eficienţă, atât în termeni cantitativi (magnitudinea efectului, de exemplu, numărul de ZFS), cât şi calitativi, prin valoarea atribuită recuperării din starea respectivă (utilitatea fiind preferinţa sau valoarea pe care o atribuie oamenii unei probleme de sănătate în funcţie de nivelul de afectare a calităţii vieţii).

Cel mai întâlnit indicator utilizat într-o analiză de tip cost-utilitate pentru a măsura eficienţa unei intervenţii este QALY (*quality adjusted life years*), acesta fiind o combinaţie între durata prelungirii vieţii şi calitatea acesteia în urma intervenţiei. De exemplu, dacă o intervenţie prelungeşte viaţa unui pacient cu 4 ani, iar nivelul calităţii vieţii pacientului aflat în condiţia respectivă este egal cu 0,50, pe o scală de la 0 (o calitate a vieţii atât de scăzută, încât moartea pare o alternativă la

fel de dezirabilă) la 1 (o calitate a vieţii oferită de o stare de sănătate perfectă), indicele este 2 QALY (4 ani × 0,50)[10].

Din nefericire, deşi analiza cost-utilitate este întâlnită frecvent în domeniul clinic, ea este aproape absentă din alte domenii de interes ale psihologiei. Însă principiul din spatele QALY, acela de a combina câştigul cantitativ (prelungirea vieţii) cu cel calitativ (trăind fericit), poate fi transpus în alte contexte. De exemplu, o intervenţie în relaţia de cuplu este interesantă nu numai dacă salvează relaţia de la divorţ, ci şi dacă contribuie la creşterea satisfacţiei maritale. În mod similar, o intervenţie din domeniul organizaţional nu este interesantă doar pentru că prelungeşte perioada de angajare în cadrul aceleiaşi firme, reducând fluctuaţia, ci şi că angajaţii rămaşi sunt mulţumiţi în cadrul firmei.

Motivul principal pentru care analizele de tip cost-utilitate au pătruns în domeniul clinic, dar deocamdată nu şi în alte domenii, derivă îndeosebi din specificul şi constrângerile financiare existente în sfera sănătăţii. Statul nu poate asigura subvenţionarea tuturor serviciilor de sănătate, iar pentru a lua raţional decizii în această direcţie, are nevoie de o unitate de măsură comună, care să permită efectuarea de comparaţii între diferite programe, chiar dacă efectul urmărit de acestea vizează o intervenţie în cancer sau rezolvarea cazurilor de arahnofobie. De aceea, metodele dezvoltate pentru a atribui o utilitate efectului urmărit sunt impregnate de modelul medical, aşa cum se poate vedea în continuare.

6.3.1. Cum se ajunge la QALY?

Analizând ideea din spatele acronimului QALY, putem vedea în acesta un produs dintre durata vieţii (LY – *life years*) şi ajustarea acestei durate în funcţie de calitatea vieţii (QA – *quality adjusted*).

LY din cadrul ecuaţiei depinde de o serie de factori, dintre care cei mai importanţi sunt: (a) tipul problemei avute în vedere (problemele de sănătate cu risc crescut sau scăzut de deces); (b) intervalul temporar avut în vedere (problemă acută *vs* cronică). De exemplu, dacă este vorba despre o intervenţie care îşi pune amprenta asupra duratei vieţii (o operaţie de extirpare a unei tumori comparativ cu absenţa intervenţiei sau cu o altă soluţie propusă spre comparare), pot exista durate

de timp diferite privind expectanţa de viaţă, situaţie în care îi revine un loc important componentei LY în calculul QALY. În asemenea cazuri se aleg intervale mai lungi de timp pentru monitorizare, astfel încât diferenţele la nivel de LY dintre intervenţii să fie reflectate în QALY (deoarece LY se referă la numărul de ani trăiţi, indiferent de calitatea vieţii din anii respectivi). În schimb, dacă intervenţia vizează o problemă care nu se reflectă în diferenţe semnificative cu privire la durata aşteptată a vieţii (de exemplu, o intervenţie menită să reducă anxietatea socială), adeseori sunt neglijabile diferenţele în termeni de LY. Într-un asemenea caz, accentul este pus pe surprinderea diferenţelor în calitatea vieţii produse de intervenţiile analizate. Această abordare este cunoscută în literatura de specialitate ca HRQoL (*Health Related Quality of Life*).

De asemenea, dacă problema este una acută, intervalul de timp de monitorizare poate fi unul scurt (preferabil cel puţin an de zile), însă pentru problemele cronice se recomandă o monitorizare îndelungată pentru calculul QALY. Aşadar, o decizie importantă în analiza cost-utilitate este alegerea intervalului temporar pentru care se calculează QALY, acesta putând varia de la un an sau chiar mai puţin (în cazul unor probleme acute, cum ar fi episoadele depresive), până la ani sau chiar zeci de ani (în cazul problemelor cronice, cum ar fi diabetul). În sfârşit, dat fiind că unitatea minimă de măsură recomandată în cazul LY este un an de zile, efectuarea de analize cost-utilitate în contextul unei perioade mai scurte de monitorizare (de exemplu, şase luni) presupune apelul la asumpţii, cum ar fi aceea că rezultatele sunt stabile şi pot fi extrapolate şi la alte perioade de timp nemonitorizate.

În privinţa celuilalt element component din QALY, QA (aspectele de calitate), acesta poate fi dedus prin metode multiple, inclusiv din datele despre eficienţă colectate în cazul unor studii de tip cost-eficienţă. În pofida diferenţelor de operaţionalizare dintre eficienţă şi utilitate, principiul este unul similar. În spatele analizei cost-utilitate se porneşte de la premisa unei legături puternice între calitatea vieţii unui individ şi simptomatologia sa. Individul are o calitate a vieţii din ce în ce mai scăzută pe măsură ce simptomatologia problemei de care suferă acesta este mai severă. De aceea, dacă se cunoaşte severitatea simptomatologiei unui individ se poate prezice nivelul calităţii vieţii sale, implicit se poate acorda o valoare utilităţii stării sale de suferinţă. De exemplu, în studiul realizat de Sava *et al.* (2009) au fost

incluşi pacienţi cu depresie majoră, care la final au obţinut niveluri diferite de eficienţă, de la menţinerea stării de depresie majoră până la recuperare completă. În funcţie de gravitatea depresiei se poate acorda o anumită valoare a utilităţii, cuprinsă, de regulă, între 0 (decedat) şi 1 (o calitate a vieţii superioară ca urmare a unei stări de sănătate fizică şi psihică perfectă). În literatura de specialitate se folosesc ponderi variabile pentru fiecare persoană ce suferă de depresie, de la 0,30 pentru cei cu depresie foarte severă până la 0,90 pentru cei recuperaţi din depresie (Bennett, Torrance, Boyle, Guscott, 2000; Freed, Rohan, Yates, 2007; Lave *et al.*, 1998; Revicki *et al.*, 1995; Revicki, Wood, 1998). De exemplu, Freed *et al.* (2007) utilizează la indivizii cuprinşi în studiu utilităţi cuprinse între 0,78 şi 0,90 pentru scoruri BDI ≤ 8; între 0,66 şi 0,78 pentru scoruri BDI între 9 şi 13; între 0,54 şi 0,66 pentru scoruri BDI cuprinse între 14 şi 19 (depresie uşoară); utilităţi între 0,42 şi 0,54 pentru depresie moderată (BDI între 20 şi 28), respectiv valori între 0,30 şi 0,42 pentru cazurile de depresie severă. Prin folosirea interpolării liniare se poate estima utilitatea pentru valori din interiorul intervalelor respective. Pornind de la asemenea sugestii, se poate transforma direct intensitatea unei probleme analizate (de exemplu, scorurile de la scala de depresie Beck) în utilităţi necesare pentru calculul QALY. Acest tip de demers este recomandat doar dacă nivelul utilităţilor nu a fost colectat de la participanţi, nici prin intermediul unor metode directe, nici prin a celor indirecte. Fiind vorba de ponderi cu care se înmulţeşte LY pentru a obţine QALY, adesea valorile utilităţilor se notează cu W (*weights*).

În studiul menţionat (Sava *et al.*, 2009) au existat date privind intensitatea depresiei colectate în trei momente de evaluare specifice: (a) pretest; (b) posttest (la trei luni distanţă de pretest) şi (c) *follow-up* (la şase luni distanţă de posttest). Transpunând formula de calcul întâlnită la ZFS pentru QALY vom avea:

QALY = [(45 de zile × Wpretest) + (135 de zile × Wposttest) + (185 de zile × W*follow-up*)]/365 zile[11].

Astfel, dacă un individ care a urmat REBT are următoarele scoruri la BDI: 28 la pretest, 13 la posttest şi 8 la finalul perioadei de urmărire, i-ar corespunde următoarele utilităţi după Freed *et al.* (2007): 0,42 la pretest, 0,54 la posttest şi 0,66 la *follow-up*. Înlocuind aceste valori în formula menţionată mai sus, am obţine 0,586 QALY, adică 214 zile de sănătate perfectă dintr-un an dacă urmează

psihoterapie REBT pentru a combate depresia majoră. În cazul QALIG (QALY *gained*), se poate analiza diferenţa în QALY dintre intervenţia activă şi un grup de control pasiv sau de tipul listei de aşteptare, iar în absenţa unor asemenea comparatori, se urmăreşte diferenţa la nivel de utilitate între posttest şi pretest zile, respectiv între *follow-up* şi pretest. Formula pentru o perioadă de un an de zile devine: QALIG = ΔW × 1 an (QALIG = (0,66 – 0,42) × 1 an). Utilizând valorile pentru cazul descris, se obţin suplimentar un număr de 87,6 ZFS sau 0,18 QALIG în urma apelului la psihoterapie REBT (luând în calcul *follow-up* şi pretestul)[12].

Prin urmare, un punct esenţial în determinarea QALY îi revine utilităţii, adică aspectului calitativ (Q) din QALY. Pentru a obţine valorile utilităţii, soluţia de lucru standard presupune colectarea empirică a utilităţilor. Astfel, metoda descrisă mai sus, aceea de a stabili utilităţi pornind de la corespondenţa dintre severitatea simptomelor măsurate şi utilităţi aşa cum apar ele în urma consultării literaturii de specialitate reprezintă doar o soluţie de avarie, atunci când nu sunt colectate direct date despre utilităţi. Pentru a colecta empiric date despre utilităţi este nevoie: (a) să se definească explicit anumite stări de sănătate; (b) să se atribuie valori numerice (utilităţi) fiecăreia dintre stările de sănătate definite, având două valori de referinţă – 0 (stare la fel de chinuitoare ca şi moartea), respectiv 1 (stare de bine socială, funcţională, psihologică şi fizică, sănătate perfectă)[13].

În măsurarea valorilor de utilitate, cel mai frecvent se apelează fie la metode directe (metoda standard a alegerii în condiţii de risc – *standard gamble approach*; metoda schimbului la timp – *time trade-off approach*; metoda vizual analogică – *visual analogic approach*), fie la metode indirecte pe bază de chestionare globale ale stării de sănătate (precum EQ-5D şi SF-GD) sau de chestionare particularizate pentru anumită tulburare în parte. Dintre toate aceste metode, chestionarele globale tind să devină din ce în ce mai răspândite, îndeosebi EQ-5D (Tolley, 2009). Ambele categorii de metode vor fi descrise succint în continuare, fiind însoţite de trimiteri bibliografice pentru cei interesaţi să aprofundeze temele prezentate. Un model didactic de studiu de tip cost-utilitate este oferit de Mitton, Rose şi Sheldon (1998).

6.3.1.1. Metodele directe de stabilire a utilităţilor

În cazul utilizării unor strategii directe de stabilire a utilităţilor, se pot pune trei întrebări importante, al căror răspuns are impact asupra valorilor finale ale utilităţilor (Tolley, 2009): (1) „ce metodă directă s-a utilizat?”; (2) „cine a oferit răspunsurile?”; (3) „au fost alese atributele adecvate pentru a aprecia şi descrie starea de sănătate?”. Caseta 6.3 ilustrează specificul celor trei tipuri majore de metode directe, cu precizarea că pentru implementarea lor există deja aplicaţii pe internet pentru obţinerea utilităţilor (vezi, de exemplu, *Health Decision Strategies*, 2012).

Caseta 6.3. *Descrierea succintă a trei metode directe de calcul ale utilităţii (Tolley, 2009)*

Alegerea în condiţii de risc

Abordarea îi solicită respondentului să opteze între două variante (două stări de sănătate): (a) una certă (de exemplu, migrene frecvente) şi (b) una posibilă, dar incertă, rezultatul putând fi, cu o anumită probabilitate, fie unul pozitiv (perfect sănătos), fie unul negativ (decedat). Utilitatea este stabilită ca fiind acea probabilitate în cazul căreia celui afectat îi vine imposibil să decidă.

Traiul cu ADHD

Dacă ar fi să trăieşti cu ADHD toată viaţa, ai accepta o intervenţie în urma căreia să existe 10% riscul să mori în urma intervenţiei şi 90 % şanse să duci o viaţă perfect sănătoasă, scăpând definitiv de toate simptomele cauzate de ADHD?

Variante de răspuns:

(A) Aş încerca noua terapie; (B) Aş rămâne cu problemele pricinuite de ADHD; (C) Mi-e imposibil (prea dificil) să mă decid.

Dacă se alege varianta C înseamnă că utilitatea acordată ADHD de persoana chestionată este de 0,90. Dacă se alege varianta B, se schimbă procentele acordate cele două stări aflate în condiţii de risc (de exemplu, 5% deces şi 95% să fie perfect sănătos) şi se reia procesul, într-o manieră iterativă, până când se alege varianta C. Principalul dezavantaj al acestei tehnici constă în tendinţa de a supraestima utilitatea unei probleme din cauza aversiunii celor mai mulţi oameni faţă de rezultate negative.

Schimbul (târgul) de timp

I se solicită individului să aleagă între două alternative, până în momentul în care îi este imposibil să decidă. În acest caz, alegerile nu se fac în condiţii de risc, ambele alternative fiind prezentate ca fiind certe.

Traiul cu ADHD

Ai fi dispus să renunţi la doi ani din viaţa ta pentru a trăi zece ani fără simptomatologie ADHD (fiind perfect sănătos)?

Variante de răspuns:

(A) Da; (B) Nu; (C) Mi-e imposibil (prea dificil) să mă decid.

Dacă se alege varianta C, utilitatea acordată ADHD de persoana chestionată este 0,80 = ((10 – 2)/10). Dacă se alege varianta B, se modifică durata anilor la care se renunţă în sensul diminuării perioadei, până se alege răspunsul C. Tehnica este destul de robustă la variaţii ale intervalului de timp propus ca stare de sănătate perfectă, în exemplul anterior acesta fiind de 10 ani (Arnesen şi Trommald, 2005).

Metoda vizual analogică

Este o metodă grafică care îi solicită celui chestionat să reprezinte grafic unde ar situa problemele sale de sănătate (să zicem ADHD) pe o scală grafică (de exemplu, cu lungimea de 10 cm), având într-un capăt decesul, iar în celălalt capăt starea perfectă de sănătate. După desenare, se măsoară unde a fost aşezată problema ADHD cu ajutorul liniarului. Bunăoară, dacă ADHD a fost reprezentat la 7 cm distanţă de punctul indicativ pentru deces, am putea afirma că utilitatea ADHD este de 0,70.

Dintre cele trei metode menţionate, alegerea în condiţii de risc tinde să ofere cele mai ridicate valori ale utilităţii, în timp ce metoda vizual analogică le oferă pe cele mai mici, iar schimbul de timp pe cele intermediare (Morimoto şi Fukui, 2002). O problemă cu impact asupra celor trei metode amintite este aceea de a stabili cine face evaluările dintre: (a) persoana suferindă; (b) specialistul; (c) publicul larg, în urma unor sondaje de opinie reprezentative, pentru a avea percepţia societăţii asupra gravităţii problemei. Nu există un răspuns clar la această problemă (Tolley, 2009), existând argumente în favoarea fiecărei variante. Totuşi, în cazul problemelor de natură psihologică, opinez că mai potrivită ar fi chestionarea populaţiei generale (comunităţii), decât a grupului-ţintă, deoarece problemele de natură psihopatologică pot impieta asupra unei analize obiective a propriei stări de sănătate. De asemenea, eşantionul investigat trebuie să fie suficient de mare (cel puţin 100) şi să fie unul eterogen pentru a asigura reprezentativitate la nivelul populaţiei generale. În cazul chestionării grupului-ţintă, important este să fie acoperită o paletă largă a problematicii respective (de exemplu, de la indivizi cu depresie uşoară până la indivizi cu depresie severă). În sfârşit, în stabilirea utilităţilor este important şi limbajul utilizat. Pot exista variaţii în utilitate dacă se prezintă generic eticheta ADHD sau dacă se oferă o descriere „dramatică” – neatent, se plictiseşte uşor, are comportamente impulsive, neliniştit şi agitat psihomotor.

6.3.1.2. Metodele indirecte de stabilire a utilităţilor

Cele mai cunoscute instrumente de acest gen sunt chestionarele care vizează starea globală de sănătate precum EQ-5D, SF-6D sau HUI (Tolley, 2009). Ele sunt mai des folosite decât metodele directe din motive de timp (durează mai puţin completarea lor) şi de fezabilitate (pot fi completate şi de către persoane care au un nivel mai scăzut de educaţie). Dintre cele trei chestionare, ghidurile NICE (NIHCE, 2012) din Marea Britanie recomandă folosirea lui EQ-5D (EuroQol – 5D) dintre metodele indirecte, urmată de utilizarea schimbului de timp dintre metodele directe. De altfel, EQ-5D este cel mai întâlnit instrument utilizat în studii de tip cost-utilitate (Whitehead şi Ali, 2010) şi cotarea unei părţi însemnate a sa are la bază metoda schimbului de timp, considerată mai acurată pentru a stabili utilităţile, decât metoda vizual analogică sau metoda alegerii în condiţii de risc (Dolan, Gudex, Kind şi Williams, 1996). EQ-5D, aşa cum este revelat din denumire, face referire la cinci dimensiuni (dificultăţi) privind: (a) mobilitatea; (b) autoîngrijirea; (c) derularea activităţilor cotidiene; (d) durerea/disconfortul şi (e) anxietatea/depresia.

Indiferent de versiunea EQ-5D, instrumentul cuprinde două părţi, care solicită individului să răspundă, în total, la şase itemi. Prima parte conţine 5 dintre ei, câte unul pentru fiecare dintre cele 5 dimensiuni. Fiecare dintre aceşti itemi conţine o descriere a stărilor de sănătate dintre care respondentul trebuie să aleagă pe aceea care i se potriveşte cel mai bine pentru dimensiunea vizată. Din combinarea celor 5 dimensiuni evaluate reiese un cod unic (de exemplu, 23321)[14], căruia i se atribuie o anumită utilitate (în acest caz de 0,15). Utilităţile sunt derivate în urma chestionării prealabile a populaţiei generale, fie prin metode statistice predictive, precum algoritmul descris în Dolan, Gudex, Kind şi Williams (1995)[15]. Cea de-a doua parte a EQ-5D este constituită pe baza modelului vizual analogic, utilitatea fiind stabilită pe baza unei reprezentări grafice oferite pentru un singur item. Scala EQ-5D poate fi consultată integral pe internet (vezi www.euroqol.org), în diferitele sale versiuni, fiind disponibilă şi în limba română, dar utilizarea ei se supune regulilor drepturilor de autor.

În anumite probleme de sănătate, îndeosebi în cele acute (de exemplu, atacul de panică) asemenea instrumente generale precum EQ-5D sau SF-6D sunt mai puţin

potrivite deoarece le lipseşte sensibilitatea (adecvarea) pentru a descrie problema respectivă. De aceea, o alternativă la scalele generice de calitate a vieţii sunt scalele specifice fiecărei tulburări, pentru a măsura problemele de calitate a vieţii strâns legate de specificul tulburării vizate. Exemple de asemenea scale sunt: Indexul Internaţional al Funcţiei Erectile (IIEF, Rosen *et al.*, 1997); Calitatea vieţii la adulţii supravieţuitori de cancer (QLACS, Avis *et al.*, 2005) ş.a.m.d. Deşi cele mai multe instrumente specifice vizează condiţii medicale, există şi unele instrumente centrate pe aspecte psihologice şi sociale, precum Scala de adaptare socială şi în muncă (WSAS, Mundt *et al.*, 2002) sau Chestionarul Calităţii Bucuriei şi Satisfacţiei în Viaţă (Q-LES-Q, Endicott *et al.*, 1993).

6.3.2. Aspecte practice în derularea unor analize cost-utilitate

Vom trece succint în revistă o serie de recomandări utile celor interesaţi să deruleze analize de tip cost-utilitate (CUA), realizând o analiză comparativă cu recomandările necesare în analizele de tip cost-eficienţă (CEA).

Tabelul 6.9. *O privire comparativă între analizele de tip cost-eficienţă şi cele de tip cost-utilitate*

Aspect urmărit	**CEA**	**CUA**	**Observaţii**
Designul studiului recomandat	Experimental	Experimental	Deşi se poate apela şi la studii cvasiexperimentale sau non-experimentale pentru a realiza CEA sau CUA, experimentele au avantajul de a avea gradul cel mai scăzut de distorsiune al efectelor măsurate prin controlul variabilelor potenţial confundate. Aceste efecte constituie, într-o formă sau alta, numitorul raportului CEA sau CUA.
Variabile de măsurat	(1) Costurile (2) VD (eficienţa)	(1) Costurile (2) VD (eficienţa) (3) Calitatea vieţii (utilitatea)	În CUA se recomandă includerea unor măsurători directe sau indirecte ale calităţii vieţii. Deşi analizele CUA pot fi efectuate şi în absenţa unor asemenea măsurători, prin apelul la transformarea valorilor VD în utilităţi în urma consultării literaturii de specialitate, aceasta reprezintă o opţiune de rezervă, atunci când nu au putut fi colectate date despre calitatea vieţii. Costurile ar trebui colectate la nivelul fiecărui participant din studiu, pe baza unor valori realiste bazate pe preţul pieţei. Pentru măsurarea eficienţei se preferă VD reale sau transformarea VD

			surogat în valori cu sens, astfel încât eficienţa să fie înţeleasă (facă sens) pentru publicul larg şi să aibă un impact social/economic/medical.
Durata studiului	Nu există vreo durată recomandată	Durata este dependentă de problema studiată	Dacă în CEA nu există vreo precizare specifică legată de durata studiului, cu excepţia cerinţei ca studiul să includă cel puţin o evaluare în etapa de *follow-up*, în CUA durata recomandată este de minimum un an de zile (ca perioadă de urmărire), dar se poate ajunge până la ani sau zeci de ani în cazul unor boli cronice.
Număr de evaluări în *follow-up*	Cât mai multe posibil	Cât mai multe posibil	În aceste evaluări accentul cade pe VD în CEA, respectiv pe utilităţi (instrumentele de calitatea vieţii) şi VD în CUA. Prin asigurarea unui număr cât mai mare de evaluări *follow-up* creşte acurateţea datelor colectate (utilităţile medii fiind estimate în intervale mai scurte de timp).
Analize posibile	Raport CE clasic, marginal şi incremental	Raport CU clasic, marginal şi incremental	Ambele analize permit diferite tipuri de raporturi (cel clasic, când se compară intervenţia cu absenţa intervenţiei; cel marginal, când se urmăreşte dacă raportul se optimizează prin extinderea sau reducerea intervenţiei; respectiv cel incremental, când se compară intervenţiile active între ele). În oricare dintre situaţiile menţionate pentru CUA, rolul central îi revine raportului cost/QALY sau raportului cost/QALYG (incremental). Analiza de sensibilitate este un procedeu obligatoriu, atât pentru CUA, cât şi pentru CEA. Spre deosebire de CEA care permite comparaţii doar dacă urmărim acelaşi efect (VD), CUA permite comparaţii între intervenţii care urmăresc efecte (VD) diferite.
Gradul de răspândire a analizei	În toate domeniile	În domeniul medical	Deşi CUA este întâlnită, în prezent, doar în sfera problemelor de sănătate (devenind treptat analiza dominantă), ideea utilităţii are o valoare euristică şi pentru alte domenii, prin faptul de a nu acorda atenţie în cadrul unui rezultat doar aspectelor cantitative (cât se prelungeşte viaţa; câţi ani de şcoală urmează; câţi ani rămâne în firmă etc.), ci şi aspectelor calitative (cât de bine se simte; cât de mulţumit este etc.).

6.4. Analiza cost-beneficiu

Analiza cost-beneficiu împărtăşeşte cu celelalte tipuri de analiză, îndeosebi cu cele de tip cost-eficienţă şi cost-utilitate, aceeaşi strategie de lucru. Astfel, se măsoară costurile, se măsoară avantajele (efectele, beneficiile), se stabileşte o perioadă de timp la care se face raportarea costurilor şi a beneficiilor, iar după obţinerea

raportului dintre costuri şi avantaje se apelează la analiza sensibilităţii pentru a urmări gradul de robusteţe a rezultatelor. Însă, spre deosebire de CEA şi CUA, analiza cost-beneficiu (CBA – *cost-benefit analysis*) se diferenţiază prin faptul că atât costurile, cât şi beneficiile sunt exprimate în termeni financiari. Dat fiind faptul că estimarea costurilor se face în acelaşi mod specificat şi pentru celelalte tipuri de evaluări economice, ne vom concentra în această secţiune asupra estimării beneficiilor în termeni financiari.

Pentru aceasta vom reveni la exemplul evaluării economice a programului de instruire adresat noilor angajaţi prezentat în secţiunea 6.2.4. În analiza cost-eficienţă costul se raporta la câştigul de tip nefinanciar – numărul de zile cu norma de muncă îndeplinită. În analiza cost-beneficiu, costurile se raportează la câştiguri estimate financiar, pentru a vedea dacă la final intervenţia aduce profit sau funcţionează pe pierdere. De aceea, aspectul esenţial în CBA şi totodată cel mai provocator din analiza cost-beneficiu este exprimarea adecvată a eficienţei intervenţiei analizate în termeni financiari. Acest lucru poate fi făcut urmând cei trei paşi descrişi în continuare, cu precizarea că estimarea este una extrem de sensibilă, astfel încât orice eroare apărută în oricare dintre paşi, distorsionează întreg rezultatul final. Ei sunt aceeaşi utilizaţi şi în cazul analizei de costuri (ce se consumă?; în ce cantităţi se consumă? ce preţ se plăteşte pentru ele?), cu precizarea că în cazul beneficiilor cele trei întrebări se ramifică pentru fiecare beneficiu în parte.

Primul pas constă în identificarea tuturor beneficiilor posibile în urma programului de training pentru noii angajaţi, cum ar fi: (a) creşterea mai rapidă a productivităţii în muncă (a performanţei celor noi instruiţi); (b) reducerea erorilor de muncă şi a defecţiunilor tehnice la aparatele pe care le manevrează; (c) reducerea fluctuaţiei de personal în stadii incipiente, pornind de la premisa că o instruire insuficientă şi lipsa supervizării şi a feedbackului constructiv constituie motive importante ale fluctuaţiei timpurii; (d) evitarea riscului de a pierde clienţi (angajaţii instruiţi fiind mai puţin supuşi riscului comiterii unor erori sau întârzierii în onorarea contractelor decât cei neinstruiţi); (e) reducerea unor costuri viitoare cu recrutarea, dacă în urma unei instruiri eficiente sunt identificaţi candidaţii potriviţi pentru procesul de recrutare internă (de exemplu, cei care asimilează cel mai rapid noile informaţii).

Al doilea pas constă în estimarea frecvenţei acestor beneficii pentru fiecare dintre sursele de câştig menţionate. De exemplu, dacă ne referim la punctul b menţionat mai sus, cel cu privire la reducerea erorilor şi a defecţiunilor tehnice, ar trebui monitorizate/evaluate aceste aspecte cât mai specific: „ce fel de defecţiuni tehnice apar, cât de des apar acestea şi ce fel de activităţi/resurse sunt necesare pentru remedierea lor?”, „necesită ele oprirea temporară a maşinilor de producţie, dacă da, pentru cât timp?”, „ce fel de piese trebuie înlocuite, în ce cantitate, sunt ele disponibile pe piaţă sau presupun perioade de aşteptare pentru livrarea lor? etc. Ar fi de dorit ca firma să posede aceste informaţii din statisticile sale oficiale şi să existe cel puţin două momente temporale de colectare a acestor informaţii: înainte de implementarea programului de instruire şi la o perioadă după instruire (pentru analize intragrup), dar poate fi şi un singur moment de evaluare postintervenţie, dacă se compară eficienţa unor programe concurente (de exemplu, instruirea la locul de muncă *vs* serviciul externalizat de trening *vs* apelul la traineri interni) (ca analiză intergrup).

Cel de-al treilea şi ultimul dintre paşii necesari este identificarea preţului corect pentru fiecare item identificat în pasul al doilea. Se iau în calcul atât costurile salvate (care au fost reduse) ca urmare a intervenţiei, cât şi veniturile obţinute în urma intervenţiei. De exemplu, dacă estimarea privind defecţiunile aparatelor indică faptul că prin introducerea instruirii *on site* a scăzut numărul de intervenţii tehnice de oprire a maşinilor automate pentru reparare la 50% din valoarea iniţială şi a scăzut cu 20% numărul rebuturilor (pieselor cu defecţiuni), aceste costuri salvate sunt considerate beneficii aduse de intervenţie. Astfel, vom avea cheltuieli de personal (ore dintr-o lună în care operatorul de bandă nu lucrează pentru că aparatul este defect, dar îşi primeşte salariul normal, costurile de manoperă cu repararea etc.); cheltuieli logistice (costul pieselor salvate de la înlocuire ca urmare a reducerii la jumătate a numărului de defecţiuni; costul la preţul pieţei al pieselor produse suplimentar ca urmare a creşterii timpului de lucru în urma reducerii timpului de staţionare). La fel se procedează cu fiecare item, acesta fiind analizat financiar, atât în termeni de costuri salvate, cât şi de venituri suplimentare aduse, pentru a se obţine o sumă finală arondată tuturor beneficiilor identificate.

După calcularea costurilor şi estimarea financiară a beneficiilor, se poate trece la analiza raportului cost-beneficiu. Una dintre cele mai cunoscute forme de analiză cost-beneficiu este returnarea investiţiei (ROI – *return of investment*). Formula ei de calcul este: totalul beneficiilor × 100%/totalul costurilor. De exemplu, dacă totalul costurilor per angajat pentru a urma un program de instruire *on site* este de 400 de dolari, iar beneficiile calculate per angajat se ridică la 1.200 de dolari, am putea spune că ROI este 300%, adică pentru fiecare dolar investit în cursul de instruire, compania câştigă trei dolari. Desigur, ca şi în cazul celorlalte tipuri de analiză, se poate apela la alte tipuri de analiză cost-beneficiu, precum CBA marginală sau CBA incrementală (acest ultim caz fiind utilizat când se compară două sau mai multe soluţii alternative)[16].

La prima impresie, analiza cost-beneficiu (CBA) prezintă o multitudine de avantaje. În primul rând foloseşte aceeaşi unitate de măsură pentru costuri şi câştiguri (avantaje, beneficii), fiind uşor de stabilit dacă costurile sunt acoperite de beneficii sau nu. În al doilea rând, beneficiile sunt foarte atrăgătoare pentru decidenţi şi în general pentru sectorul economic, deoarece sunt exprimate în termeni financiari şi fac posibilă comparaţia între programe aparţinând unor domenii diferite, ceea ce nici CEA şi nici CUA nu pot face. De exemplu, se poate compara eficienţa unui program educaţional de reducere a abandonului şcolar cu cea a unui program de screening din zona sănătăţii, cum ar cazul diabetului de tip II. În al treilea rând, spre deosebire de analiza cost-eficienţă, în care raportarea costurilor se face la un singur efect (de exemplu, se raportează eficienţa trainingului pentru noi angajaţi văzută ca număr de zile în care norma de producţie a fost îndeplinită), în analiza cost-beneficiu se încurajează urmărirea simultană a tuturor beneficiile posibile rezultate din training (de exemplu, creşterea satisfacţiei noilor angajaţi faţă de locul lor de muncă, reducerea stresului profesional ca urmare a scăderii ambiguităţii de rol etc.

În pofida acestor avantaje, analiza cost-beneficiu nu este foarte utilizată în evaluările economice ale serviciilor de psihologie, deoarece sunt greu de cuantificat financiar toate beneficiile obţinute. Dacă în CEA şi în CUA costurile reprezentă principalul aspect instabil din cadrul analizelor, astfel încât se impun analize de sensibilitate, în CBA ambii termeni ai ecuaţiei sunt extrem de instabili, deoarece

fiecare beneficiu în parte presupune propria analiză de costuri şi venituri estimate. Acest fapt conduce la multiplicarea riscului de a obţine raporturi cost-avantaje distorsionate. Bunăoară, dacă ne referim la acelaşi aspect al creşterii producţiei de piese produse ca urmare a scăderii numărului de defecţiuni tehnice printr-o mai bună manevrare a maşinilor în urma trainingului *on site*, nu putem fi siguri că ele reprezintă 100% beneficii. De exemplu, într-o piaţă foarte concurenţială, în care există probleme cu vânzarea lor, piesele suplimentare realizate rămân pe stoc, ceea ce înseamnă cheltuieli suplimentare cu depozitarea şi conservarea pieselor, costuri mai mari cu utilităţile etc. De asemenea, este posibil ca o parte dintre incidentele care au condus la defecţiuni tehnice să necesite asistenţă medicală, un cost ignorat în analiza iniţială.

Chiar dacă beneficiile au fost identificate corect (pasul 1), adesea sunt probleme privitoare la acurateţea informaţiilor de care dispune compania (de exemplu, statistici) pentru a realiza cu succes şi pasul 2. În sfârşit, rămâne mereu problema legată de măsura în care au fost estimate corect costurile (preţurile) pentru elementele identificate în primii doi paşi (pasul 3). Apoi, este posibil să fie ignorate unele costuri sau beneficii colaterale temei analizate. De exemplu, externalizarea serviciului de training poate face o impresie bună noilor angajaţi, contribuind la creşterea nivelul lor de satisfacţie în muncă şi la îmbunătăţirea climatului organizaţional. În acest context consecinţele benefice legate de creşterea productivităţii se pot propaga şi asupra celorlalţi angajaţi decât cei supuşi instruirii (efecte colaterale).

Dat fiind toate aceste limite este probabil imposibil de stabilit, fără urmă de îndoială, toate beneficiile existente în urma implementării unui program, astfel încât să avem o estimare acurată în termeni financiari a acestora. Această problemă este cu atât mai amplă, cu cât, spre deosebire de costuri, o parte dintre beneficiile stipulate nu vor fi măsurate empiric, ci estimate pe baza literaturii de specialitate şi/sau a consultării experţilor din domeniu.

Pentru a înţelege dificultatea realizării unei astfel de analize mai ales în domenii neproductive direct, cum este domeniul şcolar sau cel clinic, vom apela la schiţa propusă de Yates (1999) pentru a efectua o analiză CBA pentru un program de combatere a consumului de droguri. Autorul identifică mai multe beneficii care ar

trebui estimate financiar, ca urmare a diminuării/eliminării consumului de droguri cum ar fi: (a) reducerea cheltuielilor legale (judecată, încarcerare) ca urmare a reducerii numărului de consumatori; (b) eliminarea cheltuielilor care ar fi fost făcute pentru a cumpăra droguri; (c) reducerea cheltuielilor cu asistenţa medicală cauzate de consumul de droguri; (d) venituri suplimentare obţinute de indivizii care au renunţat la droguri ca urmare a angajării lor (prestării de activităţi economice), dar şi de stat (prin creşterea contribuţiilor sociale); (e) creşterea productivităţii în muncă, implicit a profitului angajatorului.

Acest lucru nu înseamnă că CBA nu ar trebui utilizată, ci doar că ea se potriveşte mai bine acelor situaţii care presupun un număr limitat de asumpţii. De exemplu, CBA este mai potrivită pentru a fi implementată în cazul unui training tehnic oferit agenţilor de vânzări (beneficiile fiind măsurate mai uşor prin consultarea creşterilor în vânzări observate în urma trainingului), decât în cazul unui training de dezvoltare a competenţelor de comunicare interculturală oferite profesorilor de liceu (unde beneficiile sunt mult mai greu de transpus în termeni financiari).

Concluzii

Ştiinţa a fost acuzată adesea că este prizonieră într-un turn de fildeş, fiind interesată doar de studierea (maximizarea) eficienţei intervenţiilor urmărite, dar ignorând resursele necesare şi costurile aferente pe care le implică. Pe de altă parte, deopotrivă managerii din organizaţii orientate spre profit şi decidenţii de politici publice sunt interesaţi în primul rând să găsească la problemele cu care se confruntă soluţii cu costuri rezonabile, adeseori apelând la servicii cu costul cel mai mic. Studiile de evaluare economică oferă liantul dintre cele două perspective, încercând să găsească cel mai bun raport dintre costurile pe care le presupune serviciul respectiv şi avantajele rezultate din implementarea lui. În psihologie, ca şi în cadrul altor ştiinţe cu caracter empiric, se pot realiza patru tipuri principale de evaluări economice. Analiza minimizării costurilor, cunoscută adesea sub denumirea simplă de analiză de cost, este cel mai uşor de implementat deoarece se focalizează doar pe

analiza cheltuielilor implicate. Ea conduce adesea la decizii eronate, din cauza ignorării diferenţelor dintre soluţii la nivel de beneficii aduse. Pentru a corecta acest neajuns, se poate apela la analiza cost-eficienţă şi/sau la analiza cost-utilitate. În domeniul psihologiei, analiza cost-eficienţă este cel mai întâlnit tip de evaluare economică, ea fiind secondată de analiza cost-utilitate. Indicatorii de măsurare a eficienţei în CEA sunt derivaţi din obiectivul principal al studiului, fiind posibile comparaţii între studii care urmăresc acelaşi obiectiv. Indicatorii de măsurare a eficienţei în CUA sunt standardizaţi, de obicei apelându-se la QALY, pentru a permite efectuarea de comparaţii şi între intervenţii care urmăresc obiective diferite. Însă nici CEA, nici CUA nu pot demonstra dacă beneficiile obţinute depăşesc costurile implicate, astfel încât activitatea să fie una rentabilă. Analiza cost-beneficiu este singura care poate urmări raportul dintre costuri şi beneficii în termeni de profit sau pierdere, deoarece utilizează aceeaşi unitate de măsură pentru costuri şi beneficii, fiind vorba de aspectul financiar. Pentru aceasta, CBA apelează la o multitudine de asumpţii, de a căror acurateţe depinde calitatea CBA, precum şi riscul de a obţine un raport cost-beneficiu distorsionat. De aceea, în pofida avantajelor sale, CBA este mai puţin răspândită din cauza riscurilor pe care le incumbă estimarea beneficiilor.

În sfârşit, deşi studiile de acest gen pot fi efectuate în cadrul oricărui design de cercetare, inclusiv în studii de tip non-experimental, cele experimentale maximizează şansa de a obţine raporturi acurate dintre costuri şi beneficii, ca urmare a exercitării unui control superior asupra variabilelor potenţial confundate, transpusă într-o estimare adecvată a efectelor (eficienţei unei intervenţii).

Lecturi suplimentare

Alan Haycox (2009). What is health economics? Disponibil online, găsit la adresa: http://www.whatisseries.co.uk/whatis/pdfs/What_is_health_econ.pdf

Materialul reprezintă una dintre broşurile din seria *What is…?* oferite de echipa Bandolier. Conţinutul materialului este clar, fiind trecute în revistă cele patru tipuri de analiză economică menţionate şi în acest capitol. Am ales să-l recomand datorită limbajului accesibil în care este scris, fiind util de parcurs, alături de celelalte materiale din seria respectivă referitoare la evaluările economice, fiind util mai ales acelora care au

întâmpinat dificultăţi în înţelegerea acestui capitol. Informaţii introductive despre alte concepte din sfera evaluării economice sunt disponibile la adresa: http://www.medicine.ox.ac.uk/bandolier/extra.html.

Florin A. Sava, Brian T. Yates, Viorel Lupu, Aura Szentagotai şi Daniel David (2009). Cost-Effectiveness and Cost-Utility of Cognitive Therapy, Rational Emotive Behavioral Therapy, and Fluoxetine (Prozac) in Treating Depression: A Randomized Clinical Trial. *Journal of Clinical Psychology*, 65(1), pp. 36-52.

Articolul reprezintă primul studiu românesc de analiză cost-eficienţă şi cost-utilitate din sfera psihologiei şi psihoterapiei, fiind un bun studiu de caz. Analiza cost-eficienţă respectă profilul studiilor de acest gen, reprezentând un model de urmat, poate cu excepţia modalităţii de colectare a costurilor (ar fi fost de preferat ca acestea să fie colectate individual, de la fiecare participant la studiu, nu estimate *post facto*). În schimb, analiza cost-utilitate realizată are mai multe neajunsuri cauzate îndeosebi de limitele impuse prin designul iniţial al studiului: (a) nu au fost utilizate măsurători ale calităţii vieţii pentru calculul utilităţilor; (b) perioada de urmărire a efectelor în urma intervenţiilor a fost una limitată la şase luni de zile, deşi în estimarea QALY ar fi fost preferabil să existe o perioadă de urmărire mai lungă, cu un *follow-up* de cel puţin un an de zile.

Brian T. Yates (1999). *Measuring and improving cost, cost-effectiveness, and cost-benefit for substance abuse treatment programs* [NIH Publication No. 99–4518]. Bethesda, MD: National Institute on Drug Abuse. Găsit la adresa: www.nida.nih.gov/IMPCOST/IMPCOSTIndex.html

Acest manual de lucru este axat îndeosebi pe analiza cost-eficienţă. Îl recomandăm pentru că oferă o prezentare detaliată a modului de stabilire a costurilor şi de măsurare a avantajelor obţinute în urma unei intervenţii, evidenţiind rolul central al gândirii analitice de a diseca procesele implicate pentru a permite o estimare cât mai acurată a resurselor, costurilor şi a câştigurilor implicate.

Sarah J. Whitehead şi Shehzad Ali (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, 96, pp. 5-21.

Articolul detaliază într-o manieră relativ netehnică aspectele centrale ale unei analize de tip cost-utilitate: conceptele de utilitate, respectiv QALY (numărul de ani ajustaţi calitativ). Informaţiile oferite extind discuţiile prezentate în acest capitol, fiind utile în special pentru cei care vor derula efectiv studii orientate spre analiza cost-utilitate.

Bibliografie

Arnesen, T., Trommald, M. (2005). Are QALYs based on time trade-off comparable? - A systematic review of TTO methodologies. *Health Economics*, 14(1), pp. 39-53.

Avis, N.E., Smith, K., McGraw, S., Smith, R.G., Petronis, V.M., Carver, C.S. (2005). Assessing quality of life in adult cancer survivors (QLACS). *Quality of Life Research*, 14(4), pp. 1007-1023.

Bennett, K.J., Torrance, G.W., Boyle, M.H., Guscott, R. (2000). Cost-utility analysis in depression: the McSad utility measure for depression health states. *Psychiatric Services*, 51(9), pp. 1171-1176.

David, D. (2006). *Rational treatments: Evidence-based psychological treatments for major depressive disorders*. Bucureşti: Editura Tritonic.

Denil, N., Owusu-Edusei, K., Roy, K., Schofield, A., Zohrabian, A. (s.a.). Economic Evaluation of Public Health Preparedness and Response Efforts. Găsit la adresa: http://www.cdc.gov/owcd/eet/SeriesTOC/fixed/SeriesTOC.html

Dolan, P., Gudex, C., Kind, P., Williams, A. (1995). *A social tariff for EuroQol: results from a UK general population survey.* Working Papers 138, Centre for Health Economics, University of York. Găsit la adresa: http://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE%20Discussion%20Paper%20138.pdf.

Dolan, P., Gudex, C., Kind, P., Williams, A. (1996). Valuing health states: a comparison of methods. *Journal of Health Economics*, 15(2), pp. 209-231.

Endicott, J., Nee, J., Harrison, W., Blumenthal, R. (1993). Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. *Psychopharmacology Bulletin*, 29(2), pp. 321-326.

Fals-Stewart, W., Yates, B.T., Klostermann, K. (2005). Assessing the costs, benefits, cost-benefit, and cost-effectiveness of marital and family treatments: Why we should and how we can. *Journal of Family Psychology*, 19, pp. 98-110.

Freed, M.C., Rohan, K.J., Yates, B.T. (2007). Estimating health utilities and quality adjusted life years in seasonal affective disorder research. *Journal of Affective Disorders*, 100, pp. 83-89.

Haycox, A. (2009). What is health economics? *What is…? Series.* Disponibil online, găsit la adresa: http://www.whatisseries.co.uk/whatis/pdfs/What_is_health_econ.pdf.

Health Decision Strategies (2012). [Computer softwares]. Princeton, NJ: Health Decision Strategies, LLC. Găsit la: http://www.healthstrategy.com.

Lave, J., Schulberg, H., Kamlet, M. (1998). Cost-effectiveness of treatment for depression in primary care practice. *Archives of General Psychiatry*, 55, pp. 645-651.

McCabe, C. (2009). What is cost-utility analysis? *What is…? Series.* Disponibil online, găsit la adresa: http://www.whatisseries.co.uk/whatis/pdfs/What_is_cost-util.pdf.

Mitton, C.R., Rose, M.S., Sheldon, R.S. (1998). A cost-utility analysis of pacemakers for the treatment of neurally mediated syncope. Working paper 98-5. Găsit la adresa: http://www.ihe.ca/documents/1998-05paper.pdf. Morimoto, T., Fukui, T.J (2002). Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals. *Journal of Epidemiol*ogy, 12(2), pp. 160-178.

Mundt, J.C., Marks, I.M., Shear, M.K., Greist, J.M. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry*, 180, pp. 461-464.

Phillips, C. (2009). What is cost-effectiveness? *What is…? Series.* Disponibil online, găsit la adresa: http://www.whatisseries.co.uk/whatis/pdfs/What_is_cost_effect.pdf.

Pintea, S. (2010). The relevance of results in clinical research: statistical, practical and clinical significance. *Journal of Cognitive and Behavioral Psychotherapies*, 10(1), pp. 101-114.

Revicki, D.A., Brown, R.E., Palmer, W., Bakish, D., Rosser, W.W., Anton, S.F., Feeny, D. (1995). Modeling the cost-effectiveness of antidepressant treatment in primary-care. *Pharmacoeconomics*, 8(6), pp. 524-540.

Revicki, D.A., Wood, M. (1998). Patient-assigned health state utilities for depression-related outcomes - differences by depression severity and antidepressant medications. *Journal of Affective Disorders*, 48(1), pp. 25-36.

Rosen, R.C., Riley, A., Wagner, G., Osterloh, I.H., Kirkpatrick, J., Mishra, A.(1997). The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology*, 49(6), pp. 822-830.

Sassi, F. (2006). Calculating QALYs, comparing QALY and DALY calculations. *Health Policy and Planning*, 21(5), pp. 402-408.

Sava, F.A., Yates, B.T., Lupu, V., Szentagotai, A., David, D. (2009). Cost-Effectiveness and Cost-Utility of Cognitive Therapy, Rational Emotive Behavioral Therapy, and Fluoxetine (Prozac) in Treating Depression:

A Randomized Clinical Trial. *Journal of Clinical Psychology*, 65(1), pp. 36-52.

Taylor, M. (2009). What is sensitivity analysis? *What is…? Series*. Disponibil online, găsit la adresa: http://www.whatisseries.co.uk/whatis/pdfs/What_is_sens_analy.pdf.

Tolley, K. (2009). What are health utilities? *What is…? Series*. Disponibil online, găsit la adresa: http://www.whatisseries.co.uk/whatis/pdfs/What_are_health_util.pdf.

Vannoy, S.D., Arean, P., Unützer, J. (2010). Advantages of using estimated depression-free days for evaluating treatment efficacy. *Psychiatric services*, 61(2), pp. 160-163.

Whitehead, S.J., Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, 96, pp. 5-21.

Yates, B.T. (1999). *Measuring and improving cost, cost-effectiveness, and cost-benefit for substance abuse treatment programs* (NIH Publication No. 99-4518). Disponibil online, găsit la adresa: www.nida.nih.gov/IMPCOST/IMPCOSTIndex.html.

Yates, B.T. (1994). Toward the incorporation of costs, cost-effectiveness analysis, and cost-benefit analysis into clinical research. *Journal of Consulting and Clinical Psychology*, 62, pp. 729-736.

1. În tabelul 6.2 sunt trecute direct valorile medii obţinute de la clienţii supuşi intervenţiilor respective. Pentru a creşte acurateţea datelor obţinute, în special a datelor referitoare la costurile de oportunitate, ar trebuie colectate date despre costuri într-o manieră individuală şi per şedinţă de evaluare, deoarece pot exista variaţii de la un moment la altul. Pentru aceasta, ar fi ideal ca datele despre costuri să fie colectate sub forma unui chestionar care să ofere informaţiile respective, cât mai precis. Un model de asemenea chestionar este prezentat în caseta 6.1 pentru situaţia ipotetică a unui studiu efectuat pe copii cu ADHD.

2. (scor limită maxim – scor obţinut)/(scor limită maxim – scor limită minim), în cazul BDI (29 – 10)/(29 – 7) = 0,86.

3. Se observă că aplicarea interpolării iniţiale necesită existenţa unui etalon pentru proba utilizată pentru a estima numeric nivelul VD. Dacă un asemenea etalon nu există, se poate apela la soluţii compensatorii, bazate pe conceptul de semnificaţie clinică. De pildă, în locul utilizării interpolării liniare şi a apelului la un etalon pentru a identifica scoruri-limită, Riso şi colaboratorii (1997 *apud* Vannoy, Arean şi Unützer, 2010) propun transformarea rezultatelor brute în valori dihotomice, cu scorul de 1 (dacă problema se reduce în *follow-up* la nivelul a 50% faţă de momentul iniţial al testării [*baseline*]), respectiv cu scorul de 0 dacă ameliorarea este absentă sau are o magnitudine mai redusă. Desigur, în cazul în care VD vizează o resursă care trebuie dezvoltată, poate fi utilizat un raţionament invers, scorul de 1 ar fi acordat dacă punctajul la VD în *follow-up* ar fi dublu faţă de momentul iniţial al testării. O altă variantă, mai nuanţată, deci implicit mai adecvată pentru a surprinde schimbări de o magnitudine mai scăzută, este aceea de a apela la alţi indicatori ai semnificaţiei clinice discutaţi pe scurt în lucrarea de faţă în tabelul 4.7 şi în paragrafele asociate acestuia din capitolul 4. Detalii suplimentare sunt oferite în Pintea (2010).

4. Unii autori preferă să ia în calculul ZFS doar perioada de după finalizarea intervenţiei, considerată a fi mai potrivită pentru a surprinde efectul acesteia, deoarece perioada dintre T0 şi T1 este una de dezvoltare (modificare) ce nu poate reflecta efectul integral al unei intervenţii. Atunci când perioada de *follow-up* este suficient de întinsă (de exemplu, un an de zile) această propunere poate fi una mai potrivită decât cea descrisă în caseta 6.2.

5. Pentru situaţiile în care nu se constată nicio îmbunătăţire sau se constată o înrăutăţire a situaţiei, care ar conduce la valori ale eficienţei egale cu 0 sau cu valori negative, trebuie utilizate anumite artificii tehnice, cum ar fi setarea costurilor ca şi cum ar fi existat o singură ZFS, înainte de a se apela la tehnicile statistice de comparare a diferenţelor dintre medii.

6. Această decizie putea fi intuită încă din tabelul 6.3 în care a reieşit clar că fluoxetina are un raport cost-eficienţă mai scăzut decât CT. În tabelul 6.3 se putea observa că fluoxetina era mai eficientă decât imipramina (19 ZFS *vs* 15 ZFS) şi că avea costuri mai reduse per unitate de eficienţă (21,66 *vs* 33). Astfel, imipramina are un raport cost-eficienţă chiar mai slab decât fluoxetina, care era la rândul său depăşită de CT. Pornind de la acest exemplu, se observă cât de distorsionat ar fi fost apelul la o analiză de costuri, fără a lua în considerare efectele. Ar fi fost recomandat tratamentul cu imipramină (cel mai ieftin), deşi din perspectiva raportului cost-eficienţă acest tratament ar fi avut cel mai slab raport cost-eficienţă (însemnând cele mai ridicate costuri pentru a obţine efecte similare cu celelalte tipuri de intervenţie).

7. Deoarece exemplul oferit este unul ipotetic, nu am zăbovit prea mult asupra detalierii fiecărui item din categoria costuri şi nici asupra verificării estimărilor noastre cu privire la costuri, respectiv la rezultatele (beneficiile) obţinute.

8. Costurile care privesc fiecare şedinţă de instruire pot fi privite similar consumabilelor şi raportate în întregime. Costurile care privesc simultan mai multe şedinţe de instruire (de exemplu, costul timpului alocat construirii materialelor utilizate în training) pot fi privite ca investiţii, prin urmare divizate la numărul de şedinţe alocate a avea loc (sau amortizate în timp). În sfârşit, unele costuri generale care nu vizează direct activitatea de instruire (costul videoproiectorului utilizat şi în alte activităţi) pot reprezenta un cost fracţionat (procentul din utilizarea aparatului în scopul derulării acestor programe de training, fie el închiriat sau achiziţionat, după caz) sau pot intra în categoria cheltuielilor de regie – cu procent suplimentar din totalul cheltuielilor directe).

9. Dacă programele de instruire cu scopul de integrare a noilor angajaţi ar fi urmărit un alt aspect, cum ar fi creşterea nivelului de satisfacţie al angajaţilor, atunci indicatorul bazat pe numărul de zile în care norma a fost îndeplinită ar fi fost inoportun. În asemenea cazuri, un indicator mai potrivit ar fi fost numărul de zile în care nou-angajaţii sunt mulţumiţi sau numărul de persoane al căror nivel de satisfacţie creşte în urma trainingului. Asemenea indicatori sunt greu, dar nu imposibil de operaţionalizat. Un posibil mijloc de a operaţionaliza modificările observate la nivelul satisfacţiei angajaţilor în urma programelor de instruire este indicele modificării stabile dezvoltat de Jacobson – *reliable change index* (RCI), al cărui formulă este RCI = (X post – X pre)/radical [2 × dispersie pretest × (1 – r pre-post test)]. Dacă valorile RCI sunt peste 1,96 înseamnă că satisfacţia noilor angajaţi a crescut, dacă valorile sunt mai mici de –1,96 înseamnă că satisfacţia a scăzut, iar pentru valorile cuprinse în acest interval (–1,96; +1,96), modificările la nivelul satisfacţiei în muncă sunt insuficiente pentru a vorbi despre schimbări în nivelul de satisfacţie dincolo de nivelul erorii de măsurare. În urma acestor calcule se poate urmări, spre exemplu, costul fiecărui program de instruire/nou angajat satisfăcut (cu valori ≥ 1,96). Pentru alte detalii despre RCI sau alţi indicatori similari, consultaţi Pintea (2010).

10. Formulă simplă de calcul a QALY raportată la un an de zile, fără *discount temporal* al costurilor (vezi Sassi, 2006 pentru alte variante complexe). Un alt concept standard întâlnit mai rar în analiza cost-utilitate este reversul lui QALY, cunoscut sub acronimul DALY (*disability adjusted life years*). Acesta este produsul dintre utilitatea acordată problemei respective şi costul unui an compus din zile cu simptomatologie. În acest caz, interpretarea se inversează, valori egale cu 1 înseamnă un an în condiţii de dizabilitate, iar 0 înseamnă absenţa dizabilităţii (starea de sănătate perfectă).

11. Numărul de zile alocat utilităţii din pretest decurge din faptul că această valoare contribuie pentru jumătate din perioada de urmărire dintre pretest şi posttest (cu durata de 90 de zile). În mod similar, cele 135 arondate utilităţii din posttest reprezintă însumarea a jumătate din perioada dintre pre- şi posttest (45 de zile) şi a jumătate din perioada dintre posttest şi finalul perioadei de urmărire (90 de zile din cele 180). Numărul de zile atribuite utilităţii de la finalul perioadei de urmărire reprezintă cealaltă jumătate a perioadei dintre

posttest şi *follow-up* (90 de zile), dar şi restul zilelor rămase din an, în absenţa altor date empirice colectate, pornind de la premisa că se menţine starea de sănătate observată la finalul perioadei de urmărire.

12. Dacă ar fi existat un grup de control care să nu primească o intervenţie activă, QALYG ar fi fost diferenţa dintre QALY pentru grupul cu tratament activ şi QALY pentru grupul de control pasiv (placebo, intervenţie uzuală). În studiul respectiv acest lucru nu a fost posibil deoarece toate cele trei grupe comparate vizau o intervenţie activă.

13. În practică pot exista cazuri de utilităţi negative (mai mici decât zero), dacă starea de sănătate indusă de tulburarea respectivă devine de nesuportat, fiind percepută mai rea decât decesul.

14. Spre ilustrare, codul 23321 se referă la o stare de sănătate cu anumite probleme de mers (2); incapabil să se autoîngrijească singur (să se spele sau să se îmbrace) (3); incapabil să realizeze activităţile sale curente (de muncă, studiu, timp liber) (3); cu dureri şi disconfort moderat (2); şi fără manifestări de anxietate sau depresie (1). În schimb, codul 12213 se referă la o stare de sănătate care: nu presupune niciun fel de probleme de mobilitate (1); întâmpină unele probleme de autoîngrijire (referitoare la spălat sau îmbrăcat de unul singur) (2); întâmpină unele probleme cu activităţile zilnice (muncă, studiu, timp liber etc.) (2); nu implică durere sau disconfort fizic (1); este însoţită de un nivel sever de anxietate sau depresie (3). Pentru fiecare dintre cele 5 dimensiuni sunt descrise trei stări de sănătate (funcţionare), 1 însemnând absenţa problemelor, 2 – existenţa anumitor probleme şi 3 – existenţa unor probleme grave.

15. În cazul utilizării EQ-5D valorile utilităţilor variază între 1 (perfect sănătos) şi –0,59, valoarea 0 însemnând deces. Combinaţia cea mai puţin dorită (mai puţin preferabilă decât moartea) a înregistrat utilitatea de –0,59. Ea este, desigur, combinaţia 33333, adică: incapabil să meargă (consemnat la pat); incapabil să aibă grijă de el (nu se poate spăla/îmbrăca); incapabil de a derula activităţile sale anterioare zilnice (muncă, familie, timp liber, studiu); cu dureri sau disconfort fizic intens; cu simptomatologie severă de anxietate sau depresie.

16. În acest caz al comparării a două programe active, în locul raportului beneficii × 100/costuri se utilizează valoarea netă incrementală (INPV – *incremental net present value*). INPV = (beneficiu A – beneficiu B) – (cost A – cost B), valorile pozitive însemnând o superioritate a programului A, în timp ce un rezultat cu semn minus indică o preferinţă pentru programul B.

Capitolul 7
Studiile metaanalitice

7.1. Scurt istoric şi delimitări conceptuale

Până spre mijlocul anilor ’80 ai secolului trecut, oamenii de ştiinţă se bazau exclusiv pe sinteze narative (calitative) atunci când îşi propuneau să evalueze critic un set de studii ce vizau aceeaşi problemă. De pildă, sintetizând studiile empirice din domeniul psihologiei cu privire la efectul filmelor violente asupra agresivităţii telespectatorilor, s-au obţinut rezultate contradictorii, unele studii identificând un efect semnificativ al mass-mediei în creşterea nivelului de agresivitate, altele obţinând efecte nule. Pentru a rezolva o asemenea dilemă, înainte de apariţia studiilor metaanalitice, soluţia la îndemână consta în realizarea unei evaluări calitative a studiilor (*narrative review*). Aceasta era făcută de către un expert recunoscut în domeniu, pe baza studiilor identificate pe această tematică. Problema majoră a acestui tip de abordare era faptul că nu exista o metodologie obiectivă de apreciere a studiilor, criteriile de diferenţiere dintre studii putând varia de la un cercetător la altul. De pildă, un evaluator ar fi putut considera că este nevoie de cel puţin trei studii realizate de echipe de cercetare independente pentru a considera un efect ca fiind real, în timp ce un alt cercetător era mulţumit de rezultatele unui

singur studiu pentru a indica existenţa efectului, cu condiţia ca acel studiu să fi fost bine controlat, adică să posede o validitate internă crescută.

De asemenea, în urma revoluţiei tehnologice ce a permis indexarea studiilor în baze de date internaţionale, procesul de evaluare calitativă a studiilor a devenit din ce în ce mai ineficient (întrucât solicita mari resurse de timp) şi mai puţin util (Borenstein, Hedges, Higgins şi Rothstein, 2009). Explozia de informaţii în toate domeniile de studiu este una extraordinară. De exemplu, într-un domeniu ştiinţific cum este cel al medicinei sunt publicate anual peste două milioane de articole (*apud* Mulrow, 1994).

Din aceste considerente, strategia a evoluat treptat dinspre o evaluare exclusivă de tip calitativ spre una de tip cantitativ. În acest fel s-a făcut trecerea de la evaluările narative la evaluările sistematice (*systematic review* sau *research synthesis*), ce presupun o metodologie riguroasă de lucru, în vederea diminuării subiectivităţii din procesul de evaluare.

Green şi McDonald (2005), respectiv Higgins şi Green (2011) semnalează existenţa unei neînţelegeri din partea multor specialişti, aceea de a suprapune terminologic evaluarea sistematică şi metaanaliza. Această percepţie eronată se bazează pe o situaţie reală, deoarece majoritatea evaluărilor sistematice presupun utilizarea metaanalizei ca procedură statistică de lucru, în timp ce demersul metaanalitic presupune, la rândul său, o evaluare sistematică a literaturii de specialitate. De aceea, relaţia logică corectă dintre cele două concepte este una de intersecţie. Acest lucru se întâmplă deoarece în anumite situaţii pot fi realizate studii metaanalitice pe câteva cercetări disponibile, fără a exista în prealabil o evaluare sistematică a literaturii de specialitate sau în condiţiile în care nu există suficient de multe cercetări pentru a permite extragerea unor concluzii cu privire la problema investigată. Pe de altă parte, deşi multe

evaluări sistematice se finalizează printr-o metaanaliză, există situaţii în care este nerecomandabil să fie agregate date din domenii distincte. Această critică a fost adusă chiar de unul dintre cei mai mari psihologi ai tuturor timpurilor, Hans J. Eysenck, care considera demersul metaanalitic ca fiind unul ce simplifică nepermis realitatea şi care eşuează prin tendinţa de a aduna merele cu perele (*apud* Field, 2005).

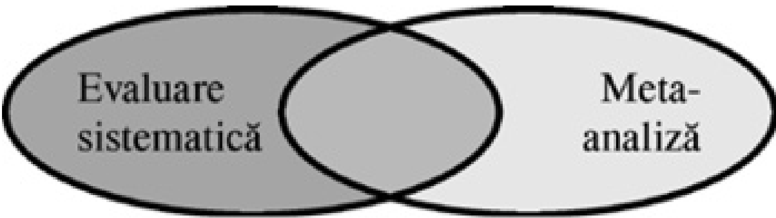


Figura 7.1. *Relaţia conceptuală dintre evaluarea sistematică şi metaanaliză*

Tabelul 7.1. *Activităţi specifice asociate conceptelor de evaluare sistematică şi metaanaliză*

Evaluarea sistematică	Metaanaliza
Definire: *Metodologia de lucru folosită pentru căutarea şi selectarea studiilor relevante (incluse în metaanaliză), cu scopul de a reduce riscul de distorsiune a rezultatelor/concluziilor*.	Definire: *Analiza statistică a datelor din studiile selectate prin procesul de evaluare sistematică, cu rolul de a creşte nivelul de obiectivitate cu privire la rezultatele/concluziile obţinute.*
Pasul 1. Formularea întrebării de investigat şi a criteriilor de includere a studiilor	Pasul 1. Calcularea indicatorilor de mărime a efectului pentru fiecare studiu eligibil
Pasul 2. Căutarea studiilor în bazele de date	Pasul 2. Selectarea tipului de metaanaliză şi a algoritmilor de calcul
Pasul 3. Completarea bazei de date prin apelul la strategii alternative de căutare	Pasul 3. Calcularea indicatorilor metaanalitici tipici (de exemplu, media ponderată, intervalul de încredere al mediei, eterogenitatea etc.)
Pasul 4. Identificarea studiilor eligibile (eliminarea studiilor neeligibile)	Pasul 4. Analize statistice suplimentare (analize pe subgrupuri de studii, metaregresii)
Pasul 5. Realizarea bazei de date şi completarea informaţiilor necesare pentru derularea procedurilor metaanalitice	Pasul 5. Verificarea robusteţii rezultatelor obţinute
Rezultat: *O bază de date cu studii eligibile ce*	Rezultat: *Un răspuns fundamentat statistic la*

conţin datele statistice necesare pentru a fi supuse metaanalizei.	*întrebarea formulată la pasul 1 al evaluării sistematice a literaturii de specialitate.*

Deşi delimitarea conceptuală menţionată mai sus este clară, se cuvine a fi precizat faptul că rezultatele celor două activităţi sunt interdependente. Astfel, fără metaanaliză sau în absenţa unei metaanalize corect efectuate, demersul de evaluare sistematică îşi pierde din însemnătate, prezentând neajunsuri similare cu cele menţionate în cazul evaluărilor narative[1], calitative. Chiar în aceste condiţii, evaluările sistematice se disting de sintezele narative prin criteriile riguroase şi transparente de selecţie a studiilor întâlnite într-un demers sistematic[2]. De asemenea, realizarea unei metaanalize în absenţa unui demers de evaluare sistematică prezintă un risc crescut de a conduce la rezultate distorsionate, deoarece numai anumite studii sunt luate în analiză. De altfel, standardele APA de publicare a metaanalizelor (vezi capitolul 8) solicită, printre altele, o detaliere a modului în care au fost selectate articolele. Fără dovada unei selecţii riguroase şi exhaustive a studiilor relevante pentru tematica investigată, metaanalizele nu au nicio şansă de a fi publicate. În consecinţă, a devenit o cutumă faptul de a combina cele două aspecte într-un demers unitar, astfel încât, atunci când se vorbeşte de demersuri metaanalitice, acestea să fie fundamentate pe o evaluare sistematică prealabilă, iar atunci când se vorbeşte de o evaluare sistematică a unui domeniu investigat, aceasta să includă o procedură metaanalitică. De aceea, pe parcursul acestui capitol vom dezvolta paşii necesari derulării unui studiu metaanalitic, pornind de la această viziune a integrării acestui demers într-unul de evaluare sistematică a problemei investigate, chiar dacă, din raţiuni pedagogice, s-a păstrat delimitarea dintre cele două aspecte teoretice.

7.2. Metodologia de evaluare sistematică. Principii şi aspecte aplicative

7.2.1. Formularea întrebării de cercetare şi a condiţiilor de eligibilitate

Confruntaţi cu numărul imens de publicaţii în orice domeniu de studiu, prima condiţie pentru eficientizarea căutărilor este să definim foarte precis ceea ce urmează să căutăm. Din perspectiva evaluării sistematice, acest lucru înseamnă deopotrivă: (a) precizarea clară a întrebării de investigat şi (b) stabilirea criteriilor de eligibilitate pentru selectarea studiilor ce fac referire la problema cercetată. Desigur, un asemenea demers ar trebui realizat doar după consultarea atentă a literaturii de specialitate, deoarece pot exista deja publicate metaanalize relevante pentru problematica investigată. Demersul poate însă continua, chiar în cazul în care se identifică metaanalize relevante pentru problematica studiată, dacă metaanalizele existente posedă o serie de limite importante din perspectiva întrebării de investigat vizate (de exemplu, includ doar studii publicate până în anul 2001 sau doar studii efectuate în SUA; pentru detalii vezi tabelul 7.2).

7.2.1.1. Formularea întrebării de investigat

În ceea ce priveşte primul aspect, cel al precizării întrebării de investigat, o variantă utilă ar fi utilizarea procedurii PICO (Spring, 2007) ce urmăreşte specificarea clară a populaţiei investigate (*population)*, a intervenţiei evaluate (*intervention*), a metodei cu rol comparativ

(*comparison*) şi a efectului urmărit (*outcome*). Acest algoritm de formulare a întrebării de investigat este deosebit de util mai ales în studiile metaanalitice de tip comparativ, în care mărimea efectului nu este reprezentată de magnitudinea coeficientului de corelaţie, ci de diferenţe între medii sau cu privire la raportul şanselor observate.

Caseta 7.1. Formulări integrale sau parţiale de tip PICO pentru întrebări de investigat

Compararea eficienţei terapiei cognitiv-comportamentale (I) cu cea a terapiei sistemice (C) în reducerea nivelului de anxietate generalizată (O) la pacienţii diagnosticaţi cu anxietate generalizată (P).

Programele de socializare bazate pe traininguri *outdoor* (I) pot reduce comportamentul antisocial (O) al adolescenţilor delincvenţi (P)*.

Psihoterapia sistemică (I) este la fel de eficientă ca şi farmacoterapia (C) în reducerea simptomatologiei depresive (O)**.

* Formulare în care lipseşte elementul comparativ
** Formulare în care populaţia nu este definită explicit

Se observă că cele mai bune formulări sunt cele mai ţintite, în care sunt precizate toate cele patru elemente. Dacă prima dintre formulări defineşte clar ceea ce este studiat, cea de-a doua nu precizează clar care este intervenţia cu care se compară eficienţa trainingurilor *outdoor*. În asemenea situaţii, lucrurile se complică, iar răspunsul la întrebare devine unul dependent de criteriile suplimentare de eligibilitate specificate. Astfel, există posibilitatea de a obţine un anumit rezultat dacă sunt selectate doar acele studii care fac referire la compararea eficienţei trainingurilor *outdoor* cu cea a grupurilor de control (de exemplu, cu evoluţia participanţilor cuprinşi în listele de aşteptare). În schimb, vom obţine un alt rezultat dacă se compară eficienţa trainingurilor *outdoor* faţă

de un grup mixt ce include ca element comparativ atât grupe de control de tipul listelor de aşteptare, cât şi grupe de control de tip placebo (intervenţii care nu sunt destinate reducerii comportamentului antisocial, dar care sunt promovate către participanţi ca metode de reducere a comportamentului antisocial).

Din cel de-al treilea exemplu prezentat în caseta 7.1 lipseşte populaţia investigată. Chiar dacă acest lucru implică, aparent, doar o extindere a studiilor eligibile, decizia poate afecta semnificativ concluzia obţinută. De pildă, este posibil ca ambele categorii de intervenţii comparate, psihoterapia şi farmacoterapia, să fie la fel de eficiente la nivel general, însă această egalitate să ascundă în spatele ei niveluri diferite de eficienţă pe categorii de vârstă (psihoterapia fiind mai eficientă la adulţi, iar farmacoterapia mai eficientă la copii şi adolescenţi). O altă problemă sesizată cu referire la cel de-al treilea exemplu este caracterul foarte general al elementului comparativ (farmacoterapia), fără să se distingă clar ce fel de tratament farmacoterapeutic se are în vedere (de exemplu, fluoxetină, imipramină etc.).

În consecinţă, ideal ar fi ca orice investigaţie sistematică să vizeze toate cele patru componente ale algoritmului PICO, chiar dacă le acordă ponderi diferite ca importanţă, în timp ce intervenţiile şi tratamentele comparative ar trebui să fie precizate cel puţin la un nivel mediu de specificitate, evitându-se concepte-umbrelă prea largi, cu un nivel ridicat de generalitate, precum farmacoterapia sau psihoterapia.

În ceea ce priveşte problemele de tip corelaţional investigate, acestea ar trebui să aibă în vedere cel puţin cele două variabile investigate (de exemplu, „asocierea dintre abilităţile cognitive generale şi performanţa la locul de muncă”; „relaţia dintre personalitate şi absenteism”), dar includerea populaţiei la care se face referire ar constitui un atu cu valenţe

aplicative ridicate. De pildă, relaţia dintre personalitate şi absenteism ar putea avea caracteristici diferite dacă se are în vedere populaţia şcolară (puberi, adolescenţi) sau populaţia activă (adulţi angajaţi), pornind de la simplul fapt că absenţele şcolare, între anumite limite, se soldează cu consecinţe instituţionale diferite comparativ cu absenţele de la locul de muncă. De multe ori însă, asemenea caracteristici ale studiilor, precum populaţia investigată, metoda de cercetare utilizată etc., nu sunt definite drept criterii de includere, ci sunt analizate ca variabile moderatoare pentru a vedea în ce contexte specifice relaţia dintre variabilele de interes (de exemplu, dintre personalitate şi absenteism) este una mai intensă. De asemenea, pot fi realizate metaanalize care să cuprindă relaţia dintre trei sau mai multe variabile de interes cum este situaţia surprinderii unor efecte de mediere (de exemplu, să studiem dacă nivelul de epuizare emoţională [*burnout*] mediază relaţia dintre neuroticism ca trăsătură de personalitate şi absenteism).

7.2.1.2. Stabilirea criteriilor de eligibilitate pentru includerea studiilor în evaluare

După stabilirea clară a problemei supuse evaluării sistematice, este nevoie să se stabilească o serie de criterii de includere (excludere) a studiilor ce abordează aceeaşi temă. Din acest punct de vedere, metodologia de evaluare sistematică a literaturii de specialitate este una riguroasă, specificată explicit în articolele de acest gen, astfel încât să fie clară cititorilor şi să permită replicarea rezultatelor.

Desigur, criteriile de eligibilitate pot urmări o serie de parametri ai studiilor, cum ar fi designul cercetării, tipologia participanţilor, modalităţile acceptate de operaţionalizare ale constructelor evaluate, data

şi limba publicării studiilor, tipul de studii etc. Cu cât sunt controlaţi mai mulţi parametri, cu atât vor rezulta mai puţine studii eligibile în final. Studiile selectate vor fi doar cele care respectă în totalitate condiţiile dorite de cercetător.

În general, evaluările sistematice care vizează doar identificarea unei asocieri între variabile (de exemplu, legătura dintre conştiinciozitate şi comportamentele contraproductive) aleg filtre mai puţin dure în selectarea studiilor, deoarece obiectivul vizat de metaanaliză este doar unul de tip descriptiv-analitic, de a identifica existenţa unei corelaţii dintre cele două constructe luând în calcul ansamblul studiilor ce conţin o asemenea analiză.

În schimb, evaluările sistematice care îşi propun obiective de tip cauzal sau explicativ utilizează criterii foarte dure în alegerea studiilor de analizat. În aceste condiţii, cercetătorul poate impune criterii dure de selecţie a studiilor, care să permită doar includerea studiilor experimentale, eliminând alte tipuri de studii, considerate problematice din perspectiva validităţii interne a unei concluzii de tip cauză-efect, precum este cazul studiilor non-experimentale. În plus, studiile metaanalitice orientate spre surprinderea unei relaţii cauzale pot lua în considerare doar anumite tipuri de experimente, în funcţie de tipul grupului comparat (de exemplu, eliminând acele studii în care grupul de control comparat nu a primit niciun fel de intervenţie).

În tabelul 7.2 sunt prezentate principalele criterii de eligibilitate utilizate în selectarea studiilor. Ele sunt însoţite de două cazuri ilustrative, adaptate din literatura de specialitate, cu referire atât la metaanaliza bazată pe comparaţii între medii (Lipsey şi Wilson, 2001), cât şi la o metaanaliză a corelaţiilor (McDaniel, 2005). Exemplele au fost modificate

în scop pedagogic pentru a reflecta cât mai precis specificul diferitelor filtre.

Tabelul 7.2. *Ilustrarea transpunerii criteriilor de eligibilitate în evaluarea sistematică*

Obiectivul evaluării sistematice	
Exemplul 1. Stabilirea eficienţei programelor *outdoor* în reducerea comportamentelor antisociale la adolescenţii delincvenţi (obiectiv cauzal). Care sunt caracteristicile celor mai eficiente, respectiv ale celor mai puţin eficiente programe (obiectiv descriptiv-analitic)? *Exemplul 2*. Stabilirea existenţei unei asocieri între nivelul de inteligenţă şi volumul creierului (obiectiv descriptiv-analitic).	
Criterii de eligibilitate: definirea acceptabilă a constructelor; tipul de studii; tipul de participanţi; contextul cultural; data publicării; limba publicării; alte filtre relevante.	
Definirea şi operaţionalizarea acceptabilă a constructelor *Scopul acestui filtru*: precizarea aspectelor definitorii pe care trebuie să le posede constructele evaluate şi a modalităţilor acceptate de operaţionalizare ale acestora. CE EVALUĂM? (validitate de construct)	*Exemplul 1*. Studiile eligibile trebuie să conţină cel puţin o intervenţie care să se încadreze ca training *outdoor* cu scopul de a reduce comportamentul antisocial al participanţilor. Prin aceasta se înţelege orice program ce implică învăţare experienţială ca urmare a derulării unor activităţi ce cuprind atât dimensiunea de provocare fizică/aventură (de exemplu, escaladarea stâncilor, deplasarea cu tiroliana etc.), cât şi dimensiunea socială (interacţiuni cu colegii, muncă în echipă, comportament prosocial). Programele de tip recreaţional (jocuri de fotbal, tenis de masă etc.) nu sunt eligibile pentru a fi incluse, deoarece le lipseşte dimensiunea de provocare fizică. Sunt, de asemenea, excluse studiile ce fac referire la programe de training *outdoor* care, deşi includ cele două componente vizate, au fost concepute în alte scopuri (de exemplu, pentru a îmbunătăţi stima de sine sau abilităţile de muncă în echipă etc.). Studiile incluse trebuie să cuprindă cel puţin o măsură validă a comportamentului antisocial printre variabilele dependente evaluate. Măsura trebuie să fie una comportamentală (de exemplu, o grilă de observaţii), nu una bazată pe răspunsuri de tip

	autoraportat (*self-report*). Variabila dependentă trebuie să facă referire directă la constructul evaluat (comportament antisocial), nefiind eligibile studiile ce vizează constructe conexe (controlul mâniei etc.). *Exemplul 2*. Studiile eligibile trebuie să includă măsurători ale ambelor constructe analizate (inteligenţa, respectiv volumul creierului). În literatura de specialitate sunt întâlnite două modalităţi de măsurare a volumului creierului: (a) măsurarea externă a circumferinţei craniului; (b) măsurarea internă, *in vivo*, a volumului creierului prin scanarea de tip RMN (imagistică cu rezonanţă magnetică). Deoarece primul tip de măsurători duce doar la o estimare a volumului creierului, vom considera eligibile pentru a fi analizate doar studiile care utilizează scanarea RMN ca modalitate de măsurare *in vivo* a volumului creierului. De asemenea, vom include doar studiile care fac referire la măsurarea integrală a volumului creierului şi vom exclude studiile care evaluează doar porţiuni ale volumului acestuia, cum ar fi evaluarea *in vivo* prin MRI a volumului de materie cenuşie frontală. În privinţa inteligenţei, vom urmări măsurarea inteligenţei fluide, aceasta fiind componenta cu cel mai mare substrat biologic, deoarece este independentă de contextul cultural.
Tipul de studii acceptate *Scopul acestui filtru*: selectarea acelor studii care pot răspunde la întrebările metaanalizei (de exemplu, a studiilor experimentale pentru obiective de tip cauzal). Prin acest criteriu se poate exercita un control asupra variabilelor străine ce pot afecta concluziile obţinute. Controlul se face prin selectarea exclusivă a studiilor care au un nivel ridicat de validitate internă, dacă obiectivul	*Exemplul 1*. Studiile luate în analiză trebuie să includă un grup comparativ. Loturile de control acceptate pot fi: tratament obişnuit (*treatment as usual*), intervenţiile de tip placebo, listele de aşteptare sau absenţa tratamentului. Elementul-cheie identificabil în grupul comparativ (de control) este absenţa efortului de a produce o schimbare în direcţia urmărită (reducerea comportamentului antisocial). Vor fi acceptate doar studiile experimentale (implică randomizarea participanţilor în grupele comparate) şi cele cvasiexperimentale (formate fără randomizarea participanţilor în grupe), cu condiţia ca pentru cele din urmă să existe

metaanalizei este unul pretenţios (de exemplu, demonstrarea unei relaţii cauzale)	evaluări la nivel de pretest (înaintea derulării intervenţiei) care să indice că loturile comparate nu diferă nici la nivelul comportamentului antisocial, nici la nivelul altor variabile asociate acestuia (care reprezintă factori de risc), cum ar fi istoricul delincvenţei, vârsta, genul participanţilor etc. Studiile de tip pretest-posttest nu sunt eligibile în absenţa unui grup de control. De asemenea, nu sunt eligibile nici studiile de tip non-experimental. *Exemplul 2*. Fiind vorba de un obiectiv de tip descriptiv-analitic, este acceptat orice tip de studiu corelaţional care include măsurători ale celor două variabile, cu condiţia respectării condiţiilor precizate la primul criteriu-filtru. Totuşi, deoarece o parte dintre probele utilizate pentru surprinderea nivelului de inteligenţă nu reprezintă
CE STUDII ACCEPTĂM PENTRU A RĂSPUNDE LA ÎNTREBAREA (OBIECTIVUL) CERCETĂRII? (validitate internă)	indicatori puri ai conceptului de inteligenţă fluidă (de exemplu, WAIS), iar altele au o validitate de construct limitată (fiind puţin răspândite/necunoscute), s-a optat pentru selectarea exclusivă a studiilor care utilizează ca măsură a inteligenţei fluide o variantă a matricilor Raven, acesta fiind instrumentul standard de evaluare a inteligenţei fluide.
Tipul de participanţi incluşi *Scopul acestui filtru*: operaţionalizarea (specificarea clară a participanţilor la care se rezumă evaluarea sistematică) CARE ESTE GRUPUL-ŢINTĂ? (validitate externă)	*Exemplul 1*. Sunt eligibile studiile realizate pe adolescenţi delincvenţi (studii în care participanţii au vârste cuprinse între 12 şi 21 de ani), indiferent de tipul infracţiunii comise şi de prezenţa sau absenţa vreunei tulburări precizate în DSM la adolescenţii incluşi în analiză (focalizarea fiind pe adolescenţii delincvenţi, în general). *Exemplul 2*. Sunt eligibili toţi participanţii, indiferent de vârstă sau sex, cu condiţia să fie vorba despre loturi neclinice, adică de participanţi care să nu fie diagnosticaţi cu vreo boală/tulburare neuropsihologică, genetică sau psihiatrică ce poate afecta datele colectate (de exemplu, capacitatea participanţilor de a completa testul de inteligenţă,

	prezenţa unor cazuri de hidrocefalie, sindrom Down etc.).
Contextul cultural *Scopul acestui filtru*: specificarea clară a gradului de generalizare a rezultatelor CARE ESTE GENERALIZABILITATEA REZULTATELOR? (validitate externă)	*Exemplul 1*. Deoarece conceptul de comportament antisocial este dependent cultural (cultura chineză, cultura din ţările islamice etc. pot avea viziuni diferite asupra măsurilor de acţiune potrivite în cazul comportamentelor antisociale), vom include în analiză doar acele studii care s-au derulat în culturi din spaţiul nord-american (SUA şi Canada). *Exemplul 2*. Sunt incluse în analiză toate studiile care respectă criteriile mai sus amintite, indiferent de spaţiul cultural în care au fost colectate datele.
Data publicării *Scopul acestui filtru*: reprezintă un filtru opţional, care poate fi activat atunci când există un motiv care ar justifica o restrângere a perioadei de căutare, cum ar fi focalizarea pe studii apărute ulterior unei metaanalize realizate, modificări în operaţionalizarea unor variabile etc. CARE ESTE PERIOADA DE RAPORTARE? (validitate externă şi de construct)	*Exemplul 1*. În analiză au fost incluse doar studii publicate după 1980. Înainte de această dată, conceptul de comportament antisocial avea o altă definire, diferită de cea actuală. Pentru a evita distorsiunile legate de ambiguităţile (diferenţele) în definirea efectului studiat (reducerea comportamentului antisocial), am luat în analiză doar studiile efectuate după 1980, care reflectă conceptul actual de comportament antisocial. *Exemplul 2*. Nu există un filtru temporal pentru includerea studiilor selectate în această evaluare sistematică. Studiile selectate iau în calcul atât cercetările incluse în analiza precedentă realizată de „Popescu" *et al.* (2003), cât şi studiile mai recente, ce nu au fost incluse în metaanaliza anterioară, mărind astfel cu 70% numărul de studii cuprinse în analiză.
Limba publicării *Scopul acestui filtru:* este un filtru opţional: barierele lingvistice pot afecta calitatea unei evaluări sistematice, mai ales atunci când nu există o explicaţie logică pentru o asemenea opţiune (vezi exemplul 2)	*Exemplul 1*. În analiză vor fi incluse doar studiile publicate în limba engleză. *Exemplul 2*. În analiză vor fi incluse studii publicate în una din următoarele limbi de circulaţie internaţională: engleză, spaniolă şi germană, pentru care există expertiza necesară de evaluare a articolelor eligibile. Pentru a elimina problematica lingvistică din această analiză, vom utiliza algoritmul de calcul de tip N sigur (*fail-safe N*)

<table>
<tr><td>CE STUDII SUNT ANALIZATE?
(validitatea externă)</td><td>pentru a identifica numărul suplimentar de studii care ar putea modifica rezultatul obţinut şi îl vom raporta la numărul de studii eligibile, dar excluse din analiză din motive lingvistice (absenţa posibilităţii de a evalua critic articolele scrise într-o limbă străină necunoscută).</td></tr>
<tr><td colspan="2">Alte filtre relevante

Tipul publicaţiei. Într-o evaluare sistematică pot fi cuprinse şi alte tipuri de studii decât cele publicate în reviste de specialitate. De exemplu, decizia de a restrânge căutările doar la nivelul studiilor publicate în reviste de specialitate poate distorsiona efectul găsit prin faptul că în reviste au mai mari şanse să pătrundă studiile cu rezultate semnificative statistic decât cele cu efecte nesemnificative statistic. De aceea, includerea rapoartelor tehnice, a prezentărilor de la conferinţe, a unor manuscrise nepublicate, a unor capitole de cărţi etc. ar oferi o viziune de ansamblu mai apropiată de realitate decât fundamentarea algoritmului de lucru doar pe baza studiilor publicate. Desigur, o asemenea decizie nu ar trebui să facă rabat de la celelalte criterii menţionate, astfel încât ar fi indicat să ne asigurăm că studiile de acest gen sunt adecvat realizate din punct de vedere metodologic, alte cauze decât calitatea lor fiind cele pentru care ele nu au ajuns să fie publicate în reviste de specialitate. O soluţie propusă de Clark-Carter (2010) ar fi efectuarea de analize distincte în situaţiile în care studiile provenite din zona gri a literaturii de specialitate (rapoarte, prezentări la conferinţe etc.) sunt incluse, respectiv excluse. Această analiză a sensibilităţii rezultatelor poate reflecta robusteţea scorurilor (în cazul în care se obţin rezultate similare), respectiv poate evidenţia fragilitatea concluziilor obţinute, în funcţie de decizia de a extinde sau nu analiza dincolo de sfera studiilor publicate în reviste de specialitate.
Datele disponibile. Unele studii, deşi eligibile din perspectiva criteriilor propuse, sunt eliminate ulterior din analize, din cauza faptului că nu conţin date suficiente pentru a identifica mărimea efectului studiat. De asemenea, alte studii pot prezenta informaţii insuficient detaliate pentru a aprecia dacă sunt sau nu îndeplinite criteriile de eligibilitate, fapt care conduce la eliminarea lor din analiză, din motive de prudenţă, mai cu seamă dacă informaţiile care lipsesc creează dificultăţi în aprecierea gradului de îndeplinire a primelor două filtre menţionate în acest tabel.</td></tr>
</table>

7.2.2. Căutarea studiilor eligibile în bazele de date existente

După stabilirea clară a întrebării de investigat (obiectivului urmărit) şi după precizarea clară a criteriilor de eligibilitate a studiilor, se poate trece

la pasul 2: căutarea studiilor relevante pentru problematica studiată. O simplă căutare folosind mijloace informatice uzuale (de exemplu, motorul de căutare Google) va da, probabil, sute sau mii de rezultate, indiferent de subiectul abordat. Pe lângă riscul obţinerii unui eşantion distorsionat şi nereprezentativ de studii, este posibil să pierdem mult timp cu identificarea studiilor relevante. Din fericire, există alte mijloace mai potrivite de a derula aceste căutări.

În primul rând se consultă bazele de date specializate pentru domeniului studiat. De pildă, în domeniul psihologiei, principala bază de date ce conţine informaţii privind studiile derulate este PsycINFO (APA, 2011). Această bază de date conţine peste 3.000.000 de intrări, dintre care 1.000.000 provin din ultimii şase ani, acoperind atât articole publicate în reviste de specialitate, cât şi capitole de cărţi, cărţi, rapoarte tehnice şi disertaţii. Deşi majoritatea înregistrărilor se rezumă la spaţiul anglo-saxon, articolele indexate sub formă de rezumate provin din reviste de specialitate publicate în peste 50 de ţări. Dezavantajul major al acestei baze de date este că înregistrează documentele doar sub formă de rezumate. Accesul la versiunea completă a acestor documente se poate face însă prin intermediul altor baze de date specializate precum PsycARTICLES (Asociaţia Psihologilor Americani, APA), Academic Search Complete/Premier (EBSCO), ProQuest Central (ProQuest), Ovid, ScienceDirect etc.

Caseta 7.2. Alte baze de date conexe domeniului psihologiei[3]

În domeniul clinic (inclusiv psihoterapie)

Medline (PubMed). Medline constituie cea mai prestigioasă bază de date din domeniul ştiinţelor medicale. PubMed este serviciul disponibil gratuit online ce indexează 5.400 de reviste din 80 de ţări, inclusiv Medline. Din păcate însă, cele mai multe înregistrări sunt disponibile doar sub formă de rezumate.

EMBASE. Este, probabil, cea mai extinsă bază de date din domeniul medical, incluzând, pe lângă înregistrările complete din Medline, alte 2.000 de reviste de specialitate neincluse în Medline, totalizând peste 24 de milioane de intrări (7.000 de reviste), la care se adaugă, din 2009, anual, rezumatele de la peste 800 de conferinţe. În schimb, are dezavantajul de a fi disponibilă contra cost.

În domeniul educaţional

ERIC. Aceasta este principala bază de date din sfera ştiinţelor educaţiei, ea incluzând articole din reviste de specialitate (aproximativ 1.200 de reviste), comunicări la conferinţe, cărţi, precum şi rapoarte tehnice din domeniul educaţiei. O parte din înregistrări este disponibilă ca text integral, însă cele mai multe înregistrări se regăsesc sub formă de rezumat.

În domeniul organizaţional

EconLit. Aceasta este principala bază de date din domeniul ştiinţelor economice, inclusiv din sfera managementului resurselor umane, cu peste un milion de înregistrări.

Interdisciplinare

Dissertation Abstracts Online/International. Această bază de date specială include aproape toate tezele de doctorat susţinute în SUA, precum şi o selecţie a tezelor de doctorat din Canada şi Marea Britanie. Din 1988 sunt incluse şi unele disertaţii de la nivel de master din SUA. Un număr considerabil de teze sunt disponibile în varianta cu text integral, nu doar sub formă de rezumate.

Scopus (SciVerse Scopus). Această bază de date combină indexarea sub formă de rezumate a literaturii de specialitate provenind din peste 16.500 de reviste din diverse domenii de studiu, dar şi a literaturii ştiinţifice identificate online prin motorul de căutare Scirus. Scopus este de asemenea, o bază de date utilă pentru a identifica citările ulterioare ale articolelor relevante pentru tema investigată.

ISI Web of Science. Este cea mai prestigioasă bază de date internaţională, reunind înregistrări din toate domeniile ştiinţifice, îndeosebi din revistele (articole, sinteze) sau conferinţele (*proceedings*) cele mai importante din fiecare domeniu de studiu. Rolul acestei baze de date este însă în primul rând unul de înregistrare a citărilor primite de fiecare articol inclus în baza de date.

Orice demers de căutare de studii în vederea realizării unei metaanalize ar trebui să pornească de la ideea căutării acestor studii în cel puţin două baze de date independente. Pentru domeniul psihologiei,

printre acestea obligatoriu ar trebui să se regăsească PsycINFO. De exemplu, studiile metaanalitice din zona intervenţiilor psihoterapeutice validate ştiinţific pentru tulburarea de depresie majoră ar putea lua în considerare trei baze de date relevante: PsycINFO, EMBASE şi Dissertation Abstracts, primele două datorită temei analizate din sfera clinică, iar ultima pentru a include acele teze de doctorat/disertaţii ce nu se regăsesc în cele două baze de date principale.

7.2.2.1. Reguli de căutare a înregistrărilor relevante pentru studiul metaanalitic

Pentru căutarea studiilor în baze de date relevante pentru tematica studiată pot fi enumeraţi câţiva algoritmi de lucru, cu scopul de a maximiza şansele de includere a cât mai multor studii relevante. Toţi algoritmii urmăresc creşterea sensibilităţii căutării în dauna specificităţii în cazul studiilor selectate. Cu alte cuvinte, se preferă mai degrabă includerea unui număr cât mai mare de studii, inclusiv a unora care se dovedesc ulterior irelevante (situaţii fals pozitive), decât includerea unui număr mai redus de studii, măsură asociată unui risc ridicat de a exclude o serie de studii relevante.

1. *Utilizaţi cuvinte-cheie incluse în tezaurul de cuvinte disponibil într-o bază de date*.

Orice bază de date conţine un tezaur de expresii (*thesaurus*), stabilite de către specialişti, ce poate fi consultat. De pildă, cunoscutul model Big Five de personalitate este precizat în tezaurul de cuvinte din PsycINFO în categoria „Five-factor model of personality”. Utilizarea cuvintelor incluse în tezaur măreşte şansa de a identifica mai multe articole relevante,

deoarece includerea lor în tezaur s-a făcut la propunerea specialiştilor pe baza frecvenţei lor de apariţie în calitate de cuvinte-cheie în articole şi a relevanţei lor conceptuale.

2. *Utilizaţi cuvinte-cheie aflate în relaţii categoriale cu conceptele-cheie studiate.*

Bunăoară, dacă cineva este interesat să studieze relaţia dintre modelul de personalitate în cinci factori şi absenteismul de la locul de muncă, căutarea nu trebuie să se rezume doar la acest nivel categorial de analiză. De exemplu, căutarea se poate extinde folosind cuvinte-cheie dintr-o categorie superioară ca nivel de abstractizare, cum ar fi personalitatea. În mod similar, în cazul absenteismului am putea utiliza conceptul de devianţă de la locul de muncă (*workplace deviance*). O asemenea strategie merită utilizată, pentru că întotdeauna vor exista studii cu niveluri diferite de generalitate (de exemplu, un studiu în care absenteismul este numai una dintre variabilele incluse în analiză poate fi parte a unui studiu extins cu privire la relaţia dintre personalitate şi comportamentul deviant la locul de muncă). În mod similar, vor exista şi studii care se vor rezuma la un nivel îngust de analiză. Un astfel de caz ar fi situaţia în care studiul ar viza doar legătura dintre conştiinciozitate şi absenteism, fără a viza toate dimensiunile modelului în cinci factori ai personalităţii. Deoarece trăsătura conştiinciozităţii este una dintre cele cinci dimensiuni ale modelului, rezultatele din asemenea studii ar fi utile pentru evaluarea relaţiei dintre dimensiunea respectivă şi absenteism.

3. *Combinaţi căutările folosind operatori logici extensivi (X sau Y).*

Pentru a maximiza probabilitatea de a cuprinde cât mai multe articole relevante pentru topica urmărită se pot utiliza cuvinte-cheie multiple, conectate prin operatorul logic de tipul „sau”. Spre ilustrare, în cazul tematicii amintite mai sus am putea utiliza o căutare care să permită

includerea oricărui studiu în analiză care conţine oricare dintre următoarele trei cuvinte-cheie: „personality” *or* „five-factor personality model” *or* „Big Five”.

4. *Utilizaţi simbolurile speciale în căutare.*

Multe baze de date, inclusiv PsycINFO, permit utilizarea unor simboluri speciale precum asteriscul (*) pentru a include toate cuvintele apropiate ca formulare lingvistică. Astfel, spre exemplu, utilizarea sintagmei „absen*” va permite identificarea articolelor care au utilizat cuvinte precum „absent”, „absence”, „absenteeism”, „absentee” etc., adică a acelor cuvinte care au aceeaşi rădăcină cu formularea ce precedă asteriscul.

5. *Faceţi apel la metacăutări folosindu-vă de istoricul căutărilor.*

Pentru a maximiza şansele căutării de articole relevante pentru tematica studiată nu este indicat ca procesul să fie finalizat într-unul sau doi paşi, ci ar fi preferabil ca această căutare să se realizeze în paşi multipli, precum cel rezumat în caseta următoare, cu privire la relaţia dintre modelul celor cinci factori de personalitate şi absenteismul la locul de muncă. Astfel, primii trei paşi descrişi au rolul de a extinde pe cât posibil aria de cuprindere a studiilor ce implică personalitatea (inclusiv, spre exemplu, modelul cu trei factori ce conţine extravertirea şi neuroticismul ca dimensiuni comune cu cele din modelul *Big Five*). Pasul 4 are, de asemenea, un rol extensiv, acela de a selecta acele studii care conţin cuvinte-cheie cu rădăcina specificată (de exemplu, „absence”, „absenteeism”, „deviant behavior”, „deviance” etc.). Rolul ultimilor doi paşi este unul de restrângere a studiilor pentru a limita căutările. Astfel, pasul 5 indică faptul că vor fi reţinute doar studiile care intersectează (reunesc) atât caracteristicile specificate în pasul 3, cât şi pe cele specificate în pasul 4. În sfârşit, ultimul pas restrânge aria de selecţie a

studiilor, păstrându-le doar pe acelea care nu se referă la absenteismul şcolar, fiind excluse studiile în care apare cuvântul-cheie „school”.

Un aspect complementar, dar important de reţinut, se referă la termenii de referinţă căutaţi. Astfel, deşi în exemplul prezentat s-a precizat peste tot că ar fi vorba despre termeni căutaţi în calitate de cuvinte-cheie, termenii pot viza şi alte specificaţii, cum ar fi faptul de a fi cuprinşi în titlu sau chiar în rezumatul înregistrării făcute. Desigur, căutarea termenilor la nivelul rezumatelor va conduce la o creştere semnificativă a studiilor care conţin termenii respectivi comparativ cu includerea acestora doar în categoria cuvintelor-cheie. De asemenea, utilizarea operatorilor logici extensivi („sau”) conduce la extinderea ariei de studii selectate, în timp ce utilizarea operatorilor logici intensivi („şi”) conduce la restrângerea studiilor selectate.

Caseta 7.3. *Paşi în căutarea de studii relevante pentru relaţia dintre personalitate şi absenteism la locul de muncă*

Pasul 1 (92548). SU personality or SU five-factor model of personality or TI big five or TI big three
Pasul 2 (12947). SU extraversion or SU conscientiousness or SU agreeableness or SU neurotic* or SU emotional stability or SU openness*
Pasul 3 (102029). Pasul 1 or Pasul 2
Pasul 4 (4562). TI absen* or TI devian* or TI attendance
Pasul 5 (213). Pasul 3 and Pasul 4
Pasul 6 (188). Pasul 5 not AB school

Observaţie! Valorile dintre paranteze reprezintă numărul de înregistrări (articole, capitole de carte etc.) găsite la momentul căutării în baza de date PsycINFO care corespund criteriului specificat. Formulările de tip SU, TI, AB pot fi specificate de la început sau se poate opta pentru scrierea directă a termenilor căutaţi, cu condiţia selectării din meniu a nivelului de analiză (titlu, cuvinte-cheie etc.). Am luat în considerare numai înregistrări care au trecut prin procesul de evaluare colegială.

În caseta 7.3 se observă că, pentru unii termeni, căutarea s-a realizat la nivelul cuvintelor-cheie (SU), pentru alţii căutarea s-a făcut la nivelul cuvintelor din titlu (TI), iar în cazul ultimului pas au fost eliminate din analiză studiile care conţineau cuvântul „school” în rezumat (AB). În final au fost identificate 188 de înregistrări ce corespund ultimului pas efectuat, ceea ce înseamnă că au fost eliminate 25 de studii faţă de pasul anterior, deoarece acestea conţineau în rezumatul lor termenul „school”, considerat irelevant pentru absenteismul la locul de muncă (al angajaţilor).

7.2.3. Strategii alternative de căutare a informaţiilor relevante

Chiar dacă apelul la bazele de date internaţionale este strategia de bază utilizată pentru a identifica studii relevante pentru tematica supusă evaluării sistematice, există, din mai multe motive, un risc crescut de a nu găsi toate documentele relevante. Un motiv ar fi acela că unele publicaţii nu ajung să fie indexate în vreo bază de date, iar altele pot fi indexate în alte baze de date decât cele utilizate în procesul de căutare. Apoi, unele rezultate ştiinţifice nu ajung să fie promovate prin articole, putând rămâne sub forma unor rapoarte de cercetare/recomandări de politici publice, realizate pentru a informa vreo agenţie publică sau vreo organizaţie privată, care adeseori este şi finanţatorul sau beneficiarul proiectului respectiv. Aceste studii fac parte din ceea ce se numeşte *literatura gri* (*grey literature*) (Hopewell, Clarke şi Mallett, 2005, *apud* Clark-Carter, 2010).

Un alt motiv pentru care unele studii relevante pentru tematica studiată nu ajung să fie identificate constă în lipsa de experienţă a cercetătorului de a alege cei mai potriviţi termeni pentru a fi introduşi în

analiza datelor. Un exemplu ar fi utilizarea exclusivă a termenului din limba engleză „absenteeism”, pentru a fi căutat în titlul înregistrărilor, spre deosebire de utilizarea criteriului mai larg de a căuta orice cuvânt din titlu care începe cu „absen*”, sau chiar a unor concepte conexe precum „workplace deviance” (vezi pasul 4 din caseta 7.3); faptul ar conduce la neselectarea unor articole potenţial relevante doar pentru faptul că în titlul lor autorii au folosit alte exprimări decât „absenteeism”.

De aceea, pentru a reduce numărul de studii relevante care nu au fost selectate după căutarea în bazele de date internaţionale cu ajutorul unor termeni-cheie se poate apela la o serie de strategii complementare, precum cele prezentate în continuare.

7.2.3.1. Identificarea unor studii din sintezele teoretice (metaanalizele) anterioare

Unele studii care au scăpat selecţiei iniţiale, în urma consultării bazelor de date internaţionale, pot fi identificate ca fiind relevante prin consultarea evaluărilor sistematice de tip narativ sau a metaanalizelor realizate pe tematica studiată, în cazul în care asemenea publicaţii există. În acest caz, se recurge atât la un demers focalizat (identificând sinteze teoretice sau metaanalize ţintite pentru tematica urmărită, cum ar fi relaţia dintre trăsăturile de personalitate după modelul celor cinci factori mari şi absenteismul de la locul de muncă), cât şi la un demers extins la tematici apropiate (sinteze teoretice sau metaanalize pe teme conexe, cum ar fi predictorii psihologici ai comportamentului deviant la locul de muncă). Printr-o asemenea strategie este posibil să identificăm alte studii relevante, datorită faptului că studiile respective erau indexate pe baza altor termeni-cheie decât cei folosiţi în algoritmul de căutare.

7.2.3.2. Identificarea unor referinţe bibliografice secundare publicate anterior

Nu numai sintezele narative/metaanalizele anterioare sunt utile pentru a identifica studii relevante. Este foarte posibil ca în urma studiului înregistrărilor relevante identificate cu ajutorul bazelor de date utilizate să găsim în textul articolelor trimiteri către referinţe bibliografice relevante, dar care nu au fost incluse între articolele selectate în primă fază.

7.2.3.3. Identificarea unor referinţe bibliografice secundare publicate ulterior

Tehnologia din ziua de astăzi, precum şi specificul unor baze de date internaţionale, îndeosebi ISI Web of Science sau Scopus, permit identificarea unor studii ulterioare publicaţiei analizate: unele dintre acestea pot fi relevante. De pildă, bazele de date amintite pot identifica toate studiile apărute (indexate) ulterior, care citează articole identificate ca relevante. Astfel, dacă din consultarea bibliografiei unui articol putem afla doar studiile similare antecedente care au stat la baza articolului respectiv, căutarea citărilor existente pentru un anumit articol permite identificarea studiilor descendente, care abordează o tematică similară cu cea a articolului evaluat.

7.2.3.4. Strategii de căutare a literaturii gri

Strategiile menţionate anterior au în vedere doar publicaţiile care, dintr-un motiv sau altul, au scăpat identificării iniţiale cu ajutorul bazei de date. Pentru a identifica cercetările nefinalizate printr-un studiu (conferinţă)

indexat în bazele de date internaţionale, se poate discuta cu specialiştii din domeniu sau se poate apela la contactarea directă a organizaţiilor relevante din domeniu. De exemplu, pentru a investiga relaţia dintre plecarea părinţilor la muncă în străinătate şi reuşita şcolară a elevilor, ar fi util să contactăm organizaţii publice sau neguvernamentale, precum Direcţiile Judeţene de Asistenţă Socială şi Protecţia Copilului, „Salvaţi Copiii” etc., care ar putea deţine studii relevante, dar nefinalizate prin publicaţii în reviste indexate în bazele de date internaţionale.

7.2.3.5. Strategii de căutare bazate pe chestionarea specialiştilor

O altă metodă utilă de căutare a unor studii suplimentare celor identificate în bazele de date internaţionale este aceea de a trimite specialiştilor interesaţi de tematica studiată o listă cu studiile identificate, cu rugămintea de a verifica dacă sunt şi alte publicaţii relevante rămase în afara listei. În caseta 7.4 este ilustrată implementarea acestei strategii în cazul unui studiu metaanalitic efectuat de McDaniel (2005) ce viza relaţia dintre inteligenţa generală şi volumul creierului.

Caseta 7.4. *Ilustrarea demersului de identificare a studiilor relevante prin apelul la specialişti*

Studiile relevante identificate prin apelul la baze de date internaţionale au fost centralizate şi trimise la peste 50 de specialişti din domeniu, fiind contactaţi: (a) cei care au publicat anterior studii privind relaţia dintre inteligenţă şi volumul creierului; (b) cei care au făcut comentarii şi analize critice pe acest subiect; (c) cei despre care se cunoştea faptul că erau interesaţi de acest subiect. Toţi aceştia au fost rugaţi să specifice eventualele studii relevante omise din lista trimisă, respectiv dacă au informaţii despre eventuale rezultate nepublicate pe această temă.

7.2.4. Identificarea studiilor eligibile pentru demersul metaanalitic

Rezultatul implementării paşilor 7.2.2 şi 7.2.3 constă, de obicei, într-un set de studii identificate ca relevante pentru tematica studiată. Însă nu toate studiile identificate sunt şi eligibile pentru a fi incluse în metaanaliză. De exemplu, unele studii nu au un design adecvat pentru a răspunde la întrebarea cercetării. Astfel, studiile de tip observaţional, non-experimental (cum ar fi demersul de stabilire a gradului de corelaţie existent între frecvenţa expunerii la jocuri video cu conţinut violent şi comportamentul agresiv al adolescenţilor) nu pot fi utilizate în metaanalize care îşi propun obiective de tip cauză-efect, menite să indice impactul jocurilor video violente asupra comportamentului agresiv la adolescenţi. Acest fapt este cauzat de raţiuni metodologice, deoarece doar studiile de tip experimental, însoţite eventual de studiile cvasiexperimentale pot răspunde la întrebări ce vizează o relaţie de tip cauză-efect (vezi capitolul 4). Desigur, acest filtru al tipului de studiu poate fi inclus încă de la început în analiză, atunci când se realizează căutarea în baza de date, însă el nu poate fi aplicat în cazul studiilor obţinute prin utilizarea strategiilor alternative de căutare. Mai mult, chiar dacă am vorbi despre studii indexate în baze de date internaţionale, apare frecvent dificultatea/imposibilitatea de a decide asupra calităţii (eligibilităţii) studiilor găsite dacă se consultă doar rezumatul articolelor relevante. Spre exemplu, dacă între criteriile de eligibilitate se specifică ideea că, dintre studiile cvasiexperimentale, vor fi luate în considerare doar acelea pentru care există dovezi privind absenţa diferenţelor dintre grupul experimental şi cel de control la nivel de pretest, este posibil ca acest lucru să nu fie precizat în sumarul articolului.

De aceea, pentru a stabili măsura în care un studiu relevant este eligibil (îndeplineşte criteriile stabilite, vezi pasul 7.2.1), trebuie să urmăm, de regulă, doi paşi: (a) screeningul rezumatelor articolelor relevante indexate în baze de date internaţionale (cu eliminarea celor care nu corespund criteriilor de eligibilitate prestabilite); (b) lecturarea articolelor care au trecut de acest prim filtru, pentru a verifica măsura în care sunt îndeplinite şi alte criterii specificate (cele care nu pot fi stabilite din rezumatul articolului), dar şi pentru a identifica alte articole relevante, ca parte a strategiei alternative de căutare a studiilor.

Cercetătorul poate decide în acest demers de analiză a eligibilităţii studiilor eliminarea celor care nu îndeplinesc anumite criterii minimale de calitate. De asemenea, cercetătorul poate decide cuantificarea calităţii studiilor considerate eligibile (acordarea de puncte) pentru a urmări impactul calităţii acestor studii asupra efectului studiat, adică măsura în care aceste diferenţe calitative dintre studii afectează rezultatul final al metaanalizei. În acest sens, vom oferi ilustrativ cazul ipotetic al unei metaanalize ce urmăreşte să evidenţieze relaţia dintre tulburarea ADHD (deficit de atenţie şi hiperchinezie) şi performanţa la şcolarii mici (ciclul primar). Un asemenea obiectiv (asocierea dintre variabilele X şi Y) permite includerea studiilor non-experimentale, alături de demersuri de tip experimental. Însă pot exista diferenţe calitative între studiile din aceeaşi categorie. De exemplu, fie studiile A şi B două studii de tip non-experimental, adică echivalente din perspectiva designului de cercetare folosit. Totuşi, studiul A poate fi considerat calitativ superior studiului B din perspectiva lotului utilizat dacă participanţii incluşi în grupul elevilor cu ADHD din studiul A sunt selectaţi în baza unui diagnostic pornind de la criteriile DSM-IV, evaluarea fiind realizată de către doi evaluatori independenţi, cu expertiză de zece ani în evaluarea psihiatrică, faţă de

cazul studiului B, în care participanţii sunt incluşi în grupul copiilor cu ADHD pe baza nominalizării făcute de cadrele didactice, cărora li s-a solicitat să numească trei elevi din fiecare clasă care manifestă dificultăţi de concentrare în timpul orelor. În mod similar, chiar dacă ambele studii ar fi selectat elevii utilizând aceeaşi metodă, cum ar fi aceea a luării în considerare a unui diagnostic recunoscut oficial, studiile ar putea diferi calitativ dacă într-unul dintre ele se specifică faptul că au fost controlate anumite variabile confundate ce pot influenţa performanţa şcolară a elevilor. Astfel, un studiu în care se precizează că nu există diferenţe la nivelul inteligenţei generale între copiii cu ADHD şi cei din lotul de control (inteligenţa generală fiind o variabilă cu impact ridicat asupra performanţei şcolare a elevilor) are un nivel metodologic calitativ superior unui studiu în care nu se specifică dacă cele două grupe de copii (ADHD şi de control) sunt echivalente din perspectiva nivelului de inteligenţă. Aşadar, chiar între studii de acelaşi nivel, cum ar fi cele non-experimentale, pot exista diferenţe calitative cauzate de anumite strategii de lucru ce pot diminua sau spori ameninţările la adresa validităţii concluziilor studiului. De aceste diferenţe ar trebui să se ţină seama în derularea unei metaanalize.

În acest sens au fost dezvoltate mai multe scale de evaluare a calităţii unui studiu, atât pentru studiile experimentale (vezi sintezele realizate de Moher şi colaboratorii, 1995; 1996), cât şi pentru studii de tip non-experimental (cum ar fi NOS – Newcastle-Ottawa Scale, dezvoltată de Wells şi colaboratorii săi, 2008). Cu toate acestea, recomandările oferite de Higgins şi Green (2011), autorii celui mai celebru ghid de evaluare sistematică (Cochrane), sunt de a nu se utiliza asemenea scale de evaluare a unui articol, finalizate prin scoruri totale. În schimb, se recomandă acordarea unui calificativ privind gradul de risc al unui articol de a avea

rezultate distorsionate (validitate internă precară). Indiferent de decizia luată, de a acorda un punctaj pentru calitatea articolului sau de a include articolul într-o categorie de risc (scăzut, ridicat sau neclar – informaţii insuficiente) pentru a cuantifica ulterior impactul includerii/excluderii acestor studii în/din analiză, principiul rămâne acelaşi. Rezultatele ar trebui însoţite de o analiză a sensibilităţii pentru a identifica relaţia existentă dintre variaţii în calitatea studiilor incluse în analiză şi magnitudinea efectului studiat.

7.2.5. Realizarea bazei de date ca input pentru metaanaliză

Realizarea bazei de date cu studiile eligibile constituie ultimul pas din procesul de evaluare sistematică. În realizarea acestui demers se redactează, pe baza întrebării de cercetare şi a criteriilor de eligibilitate definitivate în pasul 7.2.1, un chestionar de analiză pentru fiecare studiu, utilizat atât pentru realizarea secvenţei 7.2.4, privind identificarea studiilor eligibile, cât şi pentru realizarea efectivă a bazei de date, necesară pentru efectuarea analizelor statistice dorite (metaanalizei).

Structura acestui chestionar ar trebui să vizeze nu numai informaţiile statistice necesare pentru calculul indicatorilor de mărime a efectului, ci şi informaţii privind caracteristicile studiilor incluse în analiză din perspectiva metodologiei utilizate (design, participanţi, instrumente etc.), necesare pentru realizarea demersului metaanalitic.

În acest sens, deseori se construiesc chiar mai multe baze de date, unele care să vizeze caracteristicile studiilor incluse în analiză, altele care să vizeze calculul mărimii efectului pentru fiecare variabilă dependentă de interes. De asemenea, pentru a reduce riscul de codare greşită,

chestionarele sunt însoţite de un ghid de lucru (manual de codare), iar codarea efectivă are loc adesea după un scurt training (instructaj) cu rol clarificator. În plus, dacă numărul de studii eligibile este unul redus, codarea se poate realiza simultan de către doi evaluatori independenţi (realizându-se o fidelitate inter-evaluatori a codării), iar dacă numărul de studii eligibile este unul ridicat, se pot alege la întâmplare un număr de studii care să fie codate independent de un alt evaluator, pentru a verifica fidelitatea codării realizate.

Pentru a ilustra modul de colectare a datelor, vom apela la evaluarea eficienţei programelor online de instruire comparativ cu stilul clasic de instruire faţă în faţă (pentru detalii, vezi US Department of Education, 2010). Criteriile de eligibilitate şi întrebările cercetării sunt prezentate succint în tabelul 7.3, fiind urmate de o exemplificare a modului ipotetic de construire a chestionarului de colectare a datelor pe baza informaţiilor solicitate.

Tabelul 7.3. *Obiectivele studiului şi criteriile de eligibilitate. Caz ilustrativ*

Obiectivele evaluării sistematice: 1. Compararea eficienţei instruirii online *vs* faţă în faţă? 2. Ce caracteristici ale instruirii online cresc eficienţa acesteia?	
Criterii de eligibilitate	
Definirea şi operaţionalizarea acceptabilă a constructelor	Studiile eligibile includ intervenţiile de instruire realizate prin intermediul internetului, fie cu scopul de a înlocui în totalitate instruirea clasică (profesor şi elev în aceeaşi sală de clasă), fie cu scopul de a suplimenta instruirea directă, caz în care instruirea se numeşte, în limba engleză, *blended learning*. În analiză va fi cuprins orice tip de instruire online sau *blended*, fie ea una predominant expozitivă (cursantul citeşte informaţiile) sau una de tip participativ-activ (cursantul se implică activ, cum ar fi proiectele de colaborare dintre studenţi şi angajatori într-o sarcină de tip aplicativ). De asemenea, sunt luate în considerare atât programele de instruire sincrone (de exemplu,

	teleconferinţele, în care cursantul trebuie să asiste online la o dată şi o oră prestabilite, când are loc cursul), cât şi cele asincrone (programe de instruire disponibile în orice moment, nefiind necesară prezenţa cursantului online la o anumită dată/oră). Metaanaliza se va rezuma doar la acele programe disponibile pe internet, nefiind luate în calcul programe de instruire disponibile exclusiv pe CD-uri sau cele instalate local, pe un anumit calculator. Pentru măsurarea eficienţei programelor sunt incluse doar studii care evaluează obiectiv performanţa şcolară a cursanţilor (note, medii, scoruri la teste standardizate). Nu vor fi incluse în analiză studii care vizează alte rezultate, cum ar fi atractivitatea programului de instruire măsurată prin chestionare de autoraportare, modificări în preferinţele (opţiunile) studenţilor pentru disciplina studiată etc.
Tipul studiilor acceptate	Studiile eligibile sunt doar studiile experimentale sau cvasiexperimentale, cu condiţia de a supune grupul de control unui program de instruire clasic (comunicare faţă în faţă).
Felul participanţilor	Sunt incluşi participanţi de toate vârstele, de la toate nivelurile de studii, începând cu şcoala primară, indiferent de disciplina (tematica) cursului de instruire.
Contextul cultural	Sunt incluse în analiză doar programele de instruire realizate în spaţiul nord-american şi european.
Data publicării	Nu există un filtru temporal pentru includerea studiilor incluse în această evaluare sistematică.
Limba publicării	Sunt incluse doar studiile publicate în limba engleză.
Tipul publicaţiilor	Se includ toate studiile relevante identificate, atât cele publicate în fluxul principal de informaţii (de exemplu, articolele de specialitate), cât şi cele provenite din zona gri a literaturii de specialitate (de exemplu, rapoarte ale unor cercetări realizate de agenţii guvernamentale).

Caseta 7.5. *Un exemplu ipotetic de chestionar utilizat (codarea studiilor)*

Secţiunea A: IDENTIFICATORI

1. ID studiu:____

2. Autor prim: ______

3. Anul publicării: ______

4. Tipul publicaţiei
A. Carte sau capitol de carte
B. Articol
C. Teză de doctorat/Disertaţie
D. Raport tehnic
E. Comunicare la o conferinţă
F. Alt tip

Secţiunea B: PARTICIPANŢI

5. Vârsta medie _____

6. Nivelul cursului de instruire
A. Ciclul primar
B. Ciclul gimnazial
C. Ciclul liceal
D. Ciclul universitar (licenţă)
E. Ciclul universitar (master/doctorat)
F. Cursuri postuniversitare (învăţare continuă)

7. Rasa
A. > 60% albi
B. > 60% afroamericani
C. > 60% hispanici
D. eşantion mixt (nicio rasă nu depăşeşte 60%)

8. Cursanţi bărbaţi (%): ____

9. Tipul cursanţilor
A. normali
B. delincvenţi
C. cu cerinţe educative speciale (handicap)
D. cu cerinţe educative speciale (supradotaţi)

Secţiunea C: DESIGN

10. Tipul studiului
A. Non-experimental
B. Experimental
C. Cvasiexperimental cu pretest
D. Cvasiexperimental cu pretest, dar fără echivalenţa rezultatelor
E. Cvasiexperimental fără pretest
F. Neclar (nu se specifică dacă există randomizarea participanţilor)

11. Nivelul randomizării
A. Individual
B. La nivel de clasă/grup (cluster)

12. Tipul grupului de control
A. Instruire clasică (faţă în faţă)
B. Instruire online (alt model)
C. Absent

13. Nivel de retragere a participanţilor
A. Sub 10%
B. Între 10 şi 20%
C. peste 20%

14. Evaluarea globală a

designului	C. Informaţii neclare/insuficiente
A. Risc scăzut de distorsiune	
B. Risc ridicat de distorsiune	
Secţiunea D: CARACTERISTICI ALE VARIABILEI INDEPENDENTE	
15. Intervenţia	B. Instruire mixtă online + faţă în faţă (*blended learning*)
A. Instruire online (exclusivă)	
16. Stilul învăţare solicitat	
A. Pasiv (expozitiv)	C. Interactiv
B. Activ	
17. Derularea cursului online	B. Asincron
A. Sincron	
18. Tipul cursului	B. Obligatoriu de urmat (consecinţe, motivaţie extrinsecă)
A. Voluntar (intrinsec)	
19. Durata cursului	D. între 31 şi 50 de ore
A. sub 10 ore (volum de muncă)	E. între 51 şi 100 de ore
B. între 11 şi 20 de ore	F. peste 100 de ore
C. între 21 şi 30 de ore	
20. Conţinutul vizat	C. Ştiinţe umaniste/artă
A. Ştiinţe exacte şi ale vieţii	D. Mixt
B. Ştiinţe sociale	
21. Interacţiunea cu alţi colegi	
A. Posibilă faţă în faţă	D. Absentă
B. Posibilă online	E. Nespecificată
C. Posibilă mixt	
Secţiunea E: ELIGIBILITATEA VARIABILEI DEPENDENTE	
21. Tipul de rezultat vizat	
A. Notă şcolară (scor test)	C. Atitudine/preferinţă pentru subiectul studiat
B. Nivelul de abandon al cursului	D. Alt tip. Care?
C. Portofoliu de activităţi	

Suplimentar faţă de codarea datelor la nivelul studiului, de multe ori este nevoie de o codare a informaţiilor colectate din studii la nivelul variabilei dependente. Acest lucru este deseori util pentru că pot fi colectaţi mai mulţi indicatori ai variabilei dependente, care vor trebui

reuniţi ulterior într-un singur indicator compozit per studiu, pentru a nu încălca condiţia de independenţă a datelor colectate. De exemplu, dacă într-un studiu au fost colectate rezultate de la acelaşi eşantion de subiecţi cu privire la performanţele şcolare de la trei discipline (matematică, psihologie şi limba germană), acestea vor fi ulterior însumate într-un singur indicator compozit, dacă analiza realizată vizează eficienţa instruirii online în general, însă vom lua în considerare indicatorul adecvat (de exemplu, rezultatul la matematică) dacă vom analiza eficienţa instruirii online în predarea disciplinelor din domeniul ştiinţelor exacte. De aceea, pentru a putea permite asemenea diferenţieri, pe lângă codarea studiilor ce reprezintă un nivel ierarhic superior de analiză se recomandă codarea separată a indicatorilor de mărime a efectului pentru fiecare variabilă dependentă măsurată. Această codare separată vizează însă doar studiile declarate eligibile.

Caseta 7.6. *Un exemplu ipotetic de chestionar utilizat (codarea mărimii efectului)*

Secţiunea A: IDENTIFICATORI

1. ID Studiu:____

2. ID Mărimea efectului (ES): ____

Secţiunea B: CARACTERISTICI ALE VARIABILEI DEPENDENTE

3. Nivelul comparaţiei ES

A. Comparaţie pretest	C. Comparaţie *follow-up*
B. Comparaţie posttest	

4. Conţinutul disciplinei	C. Ştiinţe umaniste/artă
A. Ştiinţe exacte şi ale vieţii	D. Mixt
B. Ştiinţe sociale	

5. Sarcina evaluării notei	C. Portofoliu mixt
A. Test grilă	D. Alta. Care?
B. Proiect/Referat	

Secţiunea C: CALCULAREA MĂRIMII EFECTULUI

6. Tip de date disponibile
A. Medii, abateri standard, n grupe
B. F/t şi n grupe/grade de libertate
C. Hi pătrat
D. Frecvenţe
E. Valoarea exactă a pragului de semnificaţie
F. Alte cazuri

7. Pagina cu datele articolului: _____

8. Direcţia rezultatului
A. Online > control

9. n. coloane necesare calcul

(De exemplu, consultând caseta 7.7, în cazul diferenţelor între medii avem nevoie de valorile m, s şi n pentru grupul online, respectiv pentru grupul de control, ceea ce presupun adăugarea a şase coloane suplimentare; totuşi, mărimea efectului poate fi estimată şi pe baza altor formule, situaţie în care sunt introduse coloane suplimentare, în funcţie de datele disponibile.)

7.3. Metaanaliza

7.3.1. Calcularea indicatorilor de mărime a efectului

În privinţa modului de calcul a indicatorilor de mărime a efectului există foarte multe surse bibliografice utile, cele recomandate fiind îndeosebi lucrările realizate de Borenstein *et al.* (2009), Borenstein, Hedges, Higgins şi Rothstein (2010), Hunter şi Schmidt (1990/2004) sau Lipsey şi Wilson (2001).

De asemenea, astăzi sunt disponibile o serie de programe statistice specializate pentru calculul indicatorilor de mărime a efectului, precum Comprehensive MetaAnalysis (Borenstein *et al.*, 2011), RevMan (Cochrane Collaboration, 2011), precum şi aplicaţii specifice gratuite în programe generale precum Microsoft Excel (DeCoster, 2011), dar şi

aplicaţii de sine stătătoare (de exemplu, Becker, 1999; Lyons şi Morris, 2012).

Indiferent de sursa utilizată pentru demersul metaanalitic, se poate considera că există două mari perspective de lucru: cea propusă de Hunter şi Schmidt (1990/2004), respectiv cea propusă de Borenstein *et al.* (2009). Primii sunt autorii unui demers denumit *metaanaliza psihometrică* sau *metaanaliza componentelor esenţiale*[4] (*bare bones metaanalysis*), viziunea lor fiind mai frecvent utilizată în psihologia organizaţională şi în cea a personalităţii, îndeosebi pentru studii de tip corelaţional. Cea de-a doua propunere vine din zona ştiinţelor medicale, având un impact mai ridicat în psihologie clinică, psihoterapie, psihologie educaţională, contexte în care demersurile comparative sunt întâlnite mai frecvent. Opţiunea noastră pentru acest capitol este pentru modelul propus de Borenstein *et al.* (2009), din următoarele raţiuni: (a) modelul are un impact major asupra dezvoltării domeniului psihologiei clinice şi psihoterapiei prin implicarea în alte proiecte precum *Ghidul Cochrane de evaluare sistematică a intervenţiilor* (Higgins şi Green, 2011), în contextul în care psihologia clinică este cea mai extinsă specializare a psihologiei la nivel mondial; (b) modelul este însoţit de un software uşor de utilizat, ceea ce elimină o parte tehnică însemnată, scutindu-i pe cei interesaţi de realizarea metaanalizelor de povara calculului statistic asociat demersului metaanalitic; mai mult, la dezvoltarea acestui software au colaborat cei mai mulţi autori cu contribuţii însemnate în dezvoltarea şi popularizarea domeniul metaanalizei[5]; (c) oferă o viziune mai realistă de abordare a cazurilor în care se constată efecte eterogene; (d) constituie o abordare nouă în literatura de specialitate din România, spre deosebire de modelul propus de Hunter şi Schmidt pentru care există deja surse

bibliografice relevante (vezi metaanaliza corelaţiilor, în Maricuţoiu, 2008).

Deşi accentul este pus pe consideraţiile practice ale utilizării procedurilor metaanalitice propuse de Borenstein *et al.* (2009), vom oferi şi formulele de calcul necesare pentru demersul metaanalitic, în cazul în care nu se apelează la un software specializat. În caseta 7.7 sunt prezentate formulele necesare pentru trei cazuri diferite întâlnite în metaanalize: (a) compararea mediilor; (b) compararea frecvenţelor; (c) metaanaliza corelaţiilor.

Caseta 7.7. *Calcularea mărimii efectului după Borenstein* et al. *(2009)*

A. Cazul	comparaţiilor dintre două medii (t independent sau F având primul df = 1)
m_E, m_C	media grupului experimental, respectiv media grupului de control
s_E, s_C	abaterea standard din grupul experimental, respectiv cea din grupul de control
n_E, n_C	numărul de participanţi din grupul experimental, respectiv cel din grupul de control
t, r, N	valorile testelor t student (grupe independente); r – coeficientul de corelaţie liniară Bravais-Pearson; N – numărul total de participanţi
d, v_d, SE_d	mărimea efectului (d Cohen); dispersia mărimii efectului (v_d); eroarea standard a mărimii efectului (SE_d)
g, v_g, SE_g	mărimea efectului ajustată (g Hedges); dispersia mărimii efectului ajustate (v_g); eroarea standard a mărimii efectului ajustate (SE_g)
w	ponderea unui studiu în calculul mediei ponderate a mărimii efectului (valabilă pentru efecte fixe)

$$d = \frac{m_E - m_C}{\sqrt{\frac{(n_E - 1)s_E^2 + (n_C - 1)s_C^2}{n_E + n_C - 2}}} \quad \text{sau} \quad d = t\sqrt{\frac{n_E + n_C}{n_E n_C}} \qquad g = d\left(1 - \frac{3}{4(n_E + n_C - 2) - 1}\right)$$

$$v_d = \frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)} = SE_d^2 = \frac{1}{w} \qquad v_g = v_d\left(1 - \frac{3}{4(n_E + n_C - 2) - 1}\right)^2 = SE_g^2 = \frac{1}{w}$$

$$d = \frac{2r}{\sqrt{1 - r^2}} \qquad v_d = \frac{4\frac{(1 - r^2)^2}{N - 1}}{(1 - r^2)^3}$$

B. Cazul studiilor corelaţionale (r Bravais-Pearson)

r, N	valoarea coeficientului de corelaţie (r); volumul eşantionului
z, v_Z, SE_Z	mărimea efectului (calcul intermediar) (Z); dispersia mărimii efectului (v_Z); eroarea standard a mărimii efectului (SE_Z)
w	ponderea unui studiu în calculul mediei ponderate a mărimii efectului (valabilă pentru efect fixe)

$$z = 0,5\ln\left(\frac{1+r}{1-r}\right) \quad vs \quad r = \frac{e^{2z}-1}{e^{2z}+1} = \frac{d}{\sqrt{d^2 + \frac{(n_E+n_C)^2}{n_E n_C}}}$$

$$v_z = \frac{1}{N-3} = SE_z^2 = \frac{1}{w}$$

C. Cazul comparaţiilor dintre frecvenţe (hi pătrat având df = 1)

A, B, C, D	frecvenţe observate în cele două loturi comparate: grupul experimental (A – număr de indivizi sănătoşi; B – număr de indivizi nerecuperaţi) şi grupul de control (C – număr de indivizi sănătoşi; D – număr de indivizi nerecuperaţi)
ln OR, $v_{\ln OR}$, $SE_{\ln OR}$	logaritm natural din raportul şanselor (mărimea efectului, calcul intermediar) (lnOR); dispersia mărimii efectului (V_{lnOR}); eroarea standard a mărimii efectului (SE_{lnOR})
π, χ^2, d	pi (valoarea matematică – 3,1416); valoarea hi pătrat pentru un grad de libertate (df = 1); mărimea efectului d al lui Cohen
w	ponderea unui studiu în calculul mediei ponderate a mărimii efectului (valabilă pentru efecte fixe)

$$\ln OR = \ln\frac{AD}{BC} \quad \text{sau} \quad \ln OR = \frac{2\pi}{\sqrt{3}}\sqrt{\frac{\chi^2(1)}{N-\chi^2(1)}}$$

$$v_{\ln OR} = \frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D} = SE^2_{\ln OR} = \frac{1}{w}$$

$$d = \ln OR\frac{\sqrt{3}}{\pi} \quad v_d = v_{\ln OR}\frac{3}{\pi^2}$$

Observaţie! Unii autori utilizează notaţii inverse. De pildă DeCoster (2004), autorul Stat Help ce conţine un set de aplicaţii Excel gratuite, disponibile online pentru calculul metaanalitic, notează cu g indicatorul propus de Cohen şi cu d pe cel propus de Hedges. Noi am păstrat notaţiile d Cohen şi g Hedges pentru că sunt cele mai răspândite.

Consideraţii practice privind calculul mărimii efectului

Cele mai întâlnite modalităţi de exprimare a mărimii efectului atunci când rezultatele reprezintă comparaţia dintre două medii (grup experimental *vs* grup de control) sunt indicatorii d Cohen şi g Hedges. Ultimul ajustează valoarea lui d, deoarece aceasta tinde să fie supraestimată în cazul comparării unor grupuri cu un număr redus de participanţi incluşi în fiecare. Diferenţele dintre valorile celor doi indicatori sunt de cca 0,01. De exemplu, în cazul unei mărimi a efectului d = 0,50, g ia valoarea 0,487 pentru cazul comparării a două grupe cu câte 20 de participanţi, respectiv valoarea 0,495 pentru cazul comparării a două grupe cu câte 50 de participanţi.

Pentru a calcula mărimea efectului în contextul comparării mediilor a două grupe, este suficient să cunoaştem valorile mediilor, ale abaterilor standard şi ale numărului de persoane cuprinse în fiecare dintre cele două grupe. În absenţa acestor informaţii, mărimea efectului poate fi estimată dacă se cunoaşte valoarea lui t student rezultată din compararea celor două grupe, precum şi numărul de indivizi cuprinşi în cele două grupe. Totuşi, există o diferenţă între calcularea directă a mărimii efectului pe bază de medii, abateri standard şi număr de subiecţi faţă de estimarea indirectă a mărimii efectului pe bază de t şi N, metoda directă fiind mai exactă. De asemenea, în designuri complexe ce necesită aplicarea tehnicilor ANOVA, valorile F constituie doar aproximări ale mărimii efectului. Deşi mărimea efectului ar putea fi estimată pe baza lui F (deoarece $F = t^2$), indicatorii respectivi nu ar trebui utilizaţi în calculul metaanalizei din următoarele motive.

Dacă designul este unul unifactorial (comparaţia dintre trei sau mai multe grupe), valoarea lui F indică un test general. De exemplu, dacă se compară eficienţa terapiei cognitiv-comportamentale (TCC) *vs* terapie

familial-sistemică (TFS) *vs* medicaţie (M), ultima aleasă ca tratament standard, valoarea F nu reflectă specific relaţia dintre o intervenţie anume (TCC) şi grupul de control (M), ci reprezintă un rezultat mixt al comparării simultane a celor trei grupe. Astfel, F nu reprezintă doar comparaţia de interes pentru metaanaliză (de exemplu, între TCC şi M), ci reprezintă o valoare agregată, fiind dependentă şi de comparaţiile dintre TCC şi TFS, respectiv dintre TFS şi M.

Probleme privind valoarea lui F ca *input* în calculul mărimii efectului apar şi în cazul unor designuri factoriale cu doar două grupe de comparat per factor (de exemplu, 2 × 2). De exemplu, într-un studiu cu design intergrup 2 × 2, cei doi factori, terapie (TCC prezent *vs* absent), medicaţie (prezent *vs* absent), vor oferi trei valori F reprezentând efectul principal al terapiei, efectul principal al medicaţiei şi efectul interacţiunii dintre acestea. Chiar dacă valoarea lui F terapie ar reflecta direct efectul terapiei asupra simptomatologiei depresive prin compararea rezultatelor celor care urmează şedinţe de terapie cu ale celor care nu urmează aceste şedinţe, valoarea lui F pentru variabila independentă de interes (TCC) este dependentă de valorile celorlalte două valori F calculate (F pentru medicaţie şi F pentru efectul de interacţiune). Acest lucru apare deoarece orice valoare F reprezintă raportul dintre dispersia explicată şi dispersia rămasă neexplicată după ce au fost luate în calcul şi alte variabile, cum ar fi efectul medicaţiei sau efectul interacţiunii dintre terapie şi medicaţie. Per ansamblu, valoarea lui F (de exemplu, F pentru TCC) este dependentă de designul utilizat, tinzând să ia valori mai mari în contextul unui design cu măsurători repetate decât în cazul unui design intergrup, precum şi în contextul în care alte variabile implicate (medicaţia şi factorul de interacţiune) explică un procent mai mare din dispersia iniţial neexplicată.

O excepţie care permite utilizarea lui F pentru calculul adecvat al mărimii efectului pentru compararea evoluţiei a două grupe într-un design ANOVA este aceea a prezentării tabelului sumar ANOVA, situaţie în care F asociat comparaţiei de interes ar putea fi recalculat prin adăugarea dispersiei explicate de alţi factori la dispersia rămasă neexplicată. O altă posibilitate ar fi utilizarea rezultatelor (valorile t) obţinute la comparaţiile *post hoc* sau la testele de contrast ce vizează comparaţia de interes (de exemplu, TCC *vs* medicaţia). În sfârşit, o altă posibilitate de lucru pentru cazul unui F asociat comparării dintre două grupe obţinut într-un design complex, ar fi ajustarea mărimii efectului în funcţie de designul utilizat (vezi indicatorul eta pătrat generalizat, propus de Olejnik şi Algina [2003], urmat de convertirea datelor în termenii lui d a lui Cohen).

În cazul metaanalizelor ce includ studii corelaţionale, cea mai întâlnită modalitate de exprimare a mărimii efectului pentru cazul studiilor corelaţionale este r (chiar coeficientul de corelaţie; vezi Hunter şi Schmidt, 2004). Spre deosebire de algoritmul propus de cei doi autori, un model alternativ care a câştigat tot mai multă popularitate este cel propus de Borenstein *et al.* (2009), care solicită transformarea lui r în z Fischer pentru derularea calculului metaanalitic, urmând ca abia la final rezultatul obţinut să fie reconvertit în coeficientul de corelaţie r Bravais-Pearson. Motivul pentru care se apelează la acest subterfugiu este că în cazul utilizării lui z Fisher dispersia asociată fiecărui indicator de mărime a efectului este dependentă exclusiv de volumul eşantionului. În schimb, în cazul utilizării directe a lui r, dispersia asociată mărimii efectului este dependentă atât de volumul lotului investigat, cât şi de magnitudinea coeficientului de corelaţie.

În ştiinţele medicale îndeosebi, dar uneori şi în domeniul psihologiei este întâlnită o altă formă de metaanaliză, bazată pe studierea raportului

şanselor (*odds ratio*). Acest tip de metaanaliză este adaptat comparării programelor de intervenţie în care variabila dependentă este măsurată nominal dihotomic (de exemplu, admis/respins etc.), luându-se în calcul frecvenţele observate. De exemplu, două programe educative pot fi comparate în privinţa eficienţei lor (promovarea *vs* nepromovarea examenului de bacalaureat la prima sesiune). Două programe de consiliere familială şi de cuplu pot fi comparate ca eficienţă din perspectiva reuşitei de a evita divorţul partenerilor aflaţi în situaţii de criză (cupluri divorţate *vs* nedivorţate). Eficienţa incrementală a unui program de consiliere a carierei asociat unui curs de recalificare profesională poate fi comparată din perspectiva numărului de cursanţi şomeri reangajaţi *vs* cursanţi în continuare şomeri. În toate aceste situaţii, în care comparaţia nu se face la nivel de medii (variabilă dependentă continuă), se poate recurge la compararea frecvenţelor, pe baza raportului şanselor ca indicator al mărimii efectului. Pentru calculul mărimii efectului (raportul şanselor) trebuie să se cunoască frecvenţele pentru fiecare dintre cele patru situaţii posibile (de exemplu, cursanţi reangajaţi, respectiv cursanţi cu statutul de şomer din categoria celor care au urmat programul de consiliere în privinţa carierei, respectiv a celor care au urmat doar cursul de recalificare profesională). În lipsa acestor informaţii, raportul şanselor poate fi aproximat şi pe baza indicatorului hi pătrat (pentru un grad de libertate; vezi caseta 7.7), însă rezultatul reprezintă doar o aproximare, nefiind o transformare echivalentă. În cazul unui scor hi pătrat căruia îi corespund mai multe grade de libertate, acesta nu trebuie utilizat deoarece, fiind un indicator global (omnibuz), suferă de aceleaşi neajunsuri ca şi indicatorii F asociaţi unui design ANOVA.

În practică se întâmplă frecvent ca studiile relevante să implice indicatori diferiţi ai mărimii efectului. În cadrul casetei 7.7 sunt prezentate

modalităţi de transformare a mărimii efectului dintr-o categorie în alta (de exemplu, transformarea lui r în d Cohen, respectiv a lui ln OR în d Cohen). Trebuie menţionat faptul că aceste transformări vizează atât indicatorii de mărime a efectului în sine, cât şi valorile dispersiei asociate acestor coeficienţi.

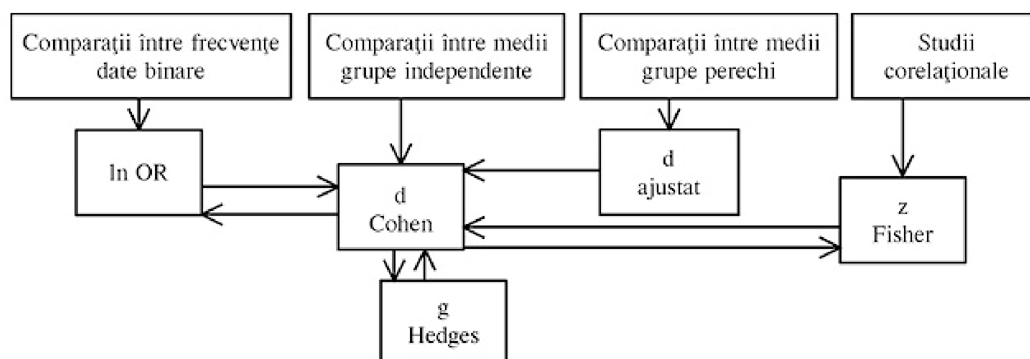


Figura 7.2. *Schema logică privind tipologia indicatorilor de mărime a efectului şi posibilităţile de convertire a acestora într-un indicator comun (adaptare după Borenstein* et al.*, 2009)*

De asemenea, se cuvine să menţionăm că, în derularea unei metaanalize, pot fi convertiţi la un numitor comun indicatori diferiţi ai mărimii efectului (de exemplu, transformarea a trei studii bazate pe ln OR în d Cohen, pentru a avea acelaşi numitor comun în toate cele 20 de studii analizate); de asemenea, este posibil să fie nevoie de convertirea la acelaşi nivel a indicatorilor de mărime a efectului proveniţi din designuri diferite (de exemplu, grupe independente *vs* grupe dependente/perechi). Indicatorii prezentaţi în cazul A din caseta 7.7 se referă la cazul cel mai frecvent întâlnit, cel al comparării a două grupe independente. Calculul mărimii efectului pentru eşantioane perechi se face pe baza altor formule (vezi Borenstein *et al.*, 2010) şi conduce la valori mai ridicate ale mărimii efectului. Pentru a putea compara metric indicatori ai mărimii efectului obţinuţi din studii cu design diferit, PowerStaTim (Maricuţoiu şi Sava, 2007) oferă un indicator d ajustat, bazat pe o formulă propusă de Morris şi

DeShon (2002), care specifică nivelul estimat al mărimii efectului în cazul în care designul ar fi fost unul de tip intergrup (grupe independente). Această facilitate unică permite includerea studiilor bazate pe un design cu măsurători repetate într-un demers metaanalitic clasic.

Nu în ultimul rând, se cuvine să facem câteva precizări cu privire la calculul indicatorilor de mărime a efectului într-o metaanaliză pentru alte situaţii decât aceea în care avem un singur indicator per studiu. De pildă, pot exista articole publicate care să conţină mai multe studii succesive, fiecare dintre acestea fiind derulat pe eşantioane diferite de participanţi. În acest caz, se poate lua fiecare studiu ca unitate independentă de analiză (de exemplu, un articol empiric alcătuit din trei studii succesive va genera trei mărimi diferite ale efectului, câte unul pentru fiecare studiu analizat). O altă situaţie posibilă este aceea în care în cadrul aceluiaşi studiu sunt incluse mai multe variabile dependente care fac referire la acelaşi concept. De exemplu, este posibil ca pentru a identifica efectul terapiei cognitiv-comportamentale asupra depresiei pe un lot de participanţi să fie utilizate două măsuri ale depresiei, scalele Beck şi Hamilton. Deoarece analiza se face pe acelaşi lot de persoane, nu este indicat să se includă în metaanaliză doi indicatori ai mărimii efectului, unul pentru scala Beck, celălalt pentru scala Hamilton, deoarece este vorba de acelaşi lot de participanţi. Soluţia corectă în astfel de cazuri este aceea a utilizării unui singur indicator per studiu, constând în media mărimii efectelor observate pentru cazul scalelor Beck, respectiv Hamilton[6]. Acestei medii îi va corespunde o dispersie ajustată, în funcţie de corelaţia dintre cele două măsurători ale aceluiaşi construct, după formula de mai jos:

$$v_{studiu} = \left(\frac{1}{m}\right)^2 \left(\sum v_i + \sum \left(r_{ij}\sqrt{v_i}\sqrt{v_j}\right)\right)$$

unde m este numărul de *outputuri* care se referă la acelaşi construct (de exemplu, m = 2 dacă sunt utilizate două scale pentru a măsura depresia); v_i, v_j sunt dispersiile asociate mărimii efectelor pentru fiecare dintre cele două scale, iar r_{ij} este corelaţia dintre cele două scale.

Aceeaşi formulă se poate aplica şi atunci când avem măsurători repetate ale unei singure variabile, cum ar fi cazul verificării efectului terapiei cognitiv-comportamentale asupra depresiei în contextul în care studiul măsoară efectul la nivel de *follow-up* pentru trei perioade distincte: la 3 luni, la 6 luni şi la 12 luni de la finalizarea terapiei.

În contextul în care nu se precizează valorile coeficienţilor de corelaţie dintre testări sau dintre variabilele output multiple, aceştia se pot estima pe baza literaturii de specialitate (preferabil) sau se pot stabili *a priori* la o valoare probabilă (de exemplu, r = 0,50).

În sfârşit, o altă situaţie practică posibilă este aceea a existenţei unor studii în care sunt comparate mai mult de două grupe. De exemplu, un studiu efectuat de David *et al.* (2008) compară eficienţa terapiei cognitiv-comportamentale clasice (CBT-Beck) cu cea a terapiei raţional-emotive (REBT-Ellis) şi cu cea a farmacoterapiei (fluoxetină) în vederea reducerii simptomatologiei depresive. Dacă ceea ce urmărim în metaanaliză este relaţia dintre REBT şi farmacoterapie, lucrurile sunt simple, deoarece se vor lua în considerare doar datele din cele două grupe. Dacă metaanaliza vizează studierea efectului terapiei cognitiv-comportamentale în general, indiferent de forma urmată, CBT clasic sau REBT, atunci va trebui utilizată o medie a mărimii efectului obţinută prin comparaţiile CBT clasic – medicaţie, respectiv REBT – medicaţie, ambele fiind relevante pentru a evalua efectul global al terapiei cognitiv-comportamentale în comparaţie cu acelaşi grup de control. Pentru calculul dispersiei asociate mediei rezultate se aplică formula menţionată mai sus, luând în

considerare o corelaţie de 0,50. O altă posibilitate de lucru, în cazul în care datele permit acest lucru, este cea de agregare a datelor provenite din cele două grupe experimentale de interes şi compararea acestui grup nou-format cu cel de control, reprezentat în acest caz de farmacoterapie.

7.3.2. Selectarea tipului de metaanaliză

După calculul manual (vezi caseta 7.7) sau automat, cu ajutorul unor aplicaţii software (Comprehensive MetaAnalysis, Borenstein *et al.*, 2011; PowerStaTim, Maricuţoiu şi Sava, 2007; Microsoft Excel – Stat Help, DeCoster, 2011 etc.), al mărimii efectului, următorul pas este de a alege modalitatea de combinare a datelor pentru a obţine rezultatele finale ale metaanalizei. Acest lucru este dependent de tipul de metaanaliză aleasă: fixă *vs* random.

Alegerea se face în funcţie de mai multe considerente, principalul fiind cel al aşteptărilor cercetătorului cu privire la cauzele diferenţelor existente de la un studiu la altul cu privire la valoarea mărimii efectului. În cazul metaanalizei efectelor fixe, cercetătorul porneşte de la premisa că există un efect teoretic egal pentru toate studiile şi că diferenţele care apar sunt cauzate de erori în eşantionare (mici variaţii de la un studiu la altul, inerente faptului că se lucrează cu eşantioane limitate ca număr de subiecţi extrase dintr-o populaţie). De exemplu, într-un studiu metaanalitic de tip corelaţional focalizat pe relaţia dintre conştiinciozitate şi performanţă în muncă, metaanaliza efectelor fixe porneşte de la ideea că variaţiile în mărimea efectului de la un studiu la altul sunt cauzate exclusiv de erori în eşantionare, cu atât mai reduse cu cât numărul persoanelor incluse în studiu este mai ridicat, deoarece efectul este acelaşi, indiferent de tipul de populaţie testată (de exemplu, manageri,

poliţişti, cadre didactice, artişti etc.). Desigur, o asemenea presupunere este greu de susţinut în practică, deoarece este foarte posibil să nu existǎ un efect similar indiferent de populaţia testatǎ, fapt ce ar presupune cǎ, pe lângǎ variaţiile determinate de erorile de eşantionare, existǎ diferenţe la nivelul mǎrimii efectului determinate de specificul populaţiei testate. De exemplu, ne aşteptǎm la un nivel diferit (mai ridicat) al corelaţiei dintre conştiinciozitate şi performanţa în muncǎ pentru cazul unor profesii cu nivel ridicat de structurare şi algoritmizare – contabili, operatori de *call center* etc.), decât în cazul unor profesii cu un nivel scǎzut de structurare (de exemplu, actori, muzicieni etc.). Aceastǎ viziune realistǎ este îmbrǎţişatǎ de modelul metaanalizei efectelor aleatorii, care porneşte de la premisa cǎ variaţiile mǎrimii efectului observate de la un studiu la altul au mai multe cauze, nu numai una. Astfel, pe lângǎ erorile de eşantionare (dispersie intrastudiu) prezumate şi în cazul metaanalizei efectelor fixe, apar variaţii determinate de diferenţele calitative dintre populaţiile din care au fost extrase eşantioanele de participanţi la studii. Cu alte cuvinte, nu se porneşte de la premisa cǎ ar exista un efect fix teoretic, ci cǎ aceste efecte sunt distribuite normal în jurul unei medii, fiind dependente şi de variaţiile sistematice de la nivelul studiilor (dispersie interstudii).

Din cele prezentate rezultǎ cǎ metaanaliza efectelor random porneşte de la o premisǎ mai realistǎ decât metaanaliza efectelor fixe, deoarece este puţin probabil sǎ gǎsim un efect fix indiferent de populaţia investigatǎ. Cu toate acestea, destul de mulţi autori apeleazǎ la metaanaliza efectelor fixe, deoarece: (a) este mai puţin pretenţioasǎ din perspectiva numǎrului de studii putând fi efectuatǎ chiar şi în situaţia reunirii unui numǎr redus de studii (de exemplu, 10-15), în timp ce metaanaliza efectelor aleatorii necesitǎ un numǎr ridicat de studii (în general peste 20) pentru a oferi indicatori statistici nedistorsionaţi[7]; (b) este mai puţin conservatoare,

puterea statistică fiind mai ridicată într-o metaanaliză a efectelor fixe decât într-o metaanaliză a efectelor random, prin urmare fiind şanse mai mari de a găsi efecte semnificative statistic într-o metaanaliză a efectelor fixe; (c) este mai uşor de realizat dacă se apelează la calculul manual, deoarece implică formule mai puţin complicate (vezi comparativ casetele 7.7. şi 7.8 privind coeficienţii de ponderare w asociaţi metaanalizei efectelor fixe, respectiv efectelor random după Borenstein *et al.*, 2009).

Totuşi, cu excepţia cazului în care există un număr redus de studii, recomandarea noastră ar fi aceea de a utiliza metaanaliza efectelor aleatorii datorită premiselor realiste de la care porneşte acest gen de analiză şi datorită filtrelor suplimentare utilizate (de exemplu, nu se ţine cont în calcularea efectului final mediu doar de variaţii în volumul eşantioanelor, ci şi de gradul de omogenitate a indicatorilor obţinuţi, penalizând astfel efectele extreme, adică situaţiile în care rezultatele obţinute sunt mai puţin consistente). În sfârşit, o altă raţiune semnificativă pentru a utiliza metaanaliza efectelor aleatorii ori de câte ori este posibil are în vedere concluziile ce pot fi desprinse din metaanaliză. În metaanaliza efectelor fixe rezultatul obţinut nu poate fi extrapolat, el fiind valabil doar pentru tipul de indivizi investigaţi. În metaanaliza efectelor aleatorii rezultatul final obţinut are un nivel ridicat de generalizare, inclusiv la alte populaţii decât cele incluse în studiu.

În sfârşit, o situaţie particulară este realizarea unei metaanalize a efectelor mixte, posibilă deoarece într-un demers metaanalitic pot fi urmărite mai multe niveluri ierarhice (de exemplu, efectul global pentru relaţia dintre conştiinciozitate şi performanţa în muncă – un nivel ierarhic superior –, cât şi o serie de efecte valabile la nivel de subgrupe de studiu, cum ar fi relaţia dintre conştiinciozitate şi performanţa în muncă în trei situaţii distincte: pentru funcţii manageriale [de decizie]; pentru funcţii

executive slab structurate [antreprenori, artişti] şi pentru funcţii executive puternic structurate [contabili, operatori pe maşini automatizate etc.]). Dacă se aplică modele de lucru diferite – de exemplu, metaanaliza efectelor aleatorii în cadrul fiecăruia dintre cele trei domenii exemplificate, respectiv metaanaliza efectelor fixe pentru rezultatul global –, vorbim despre cazul special al metaanalizei efectelor mixte.

Implicaţii practice ale alegerii

O implicaţie imediată constă în algoritmul de calculare a mediei ponderate a mărimii efectelor din studiile analizate. Astfel, în niciunul dintre tipurile de metaanaliză menţionate media calculată nu se rezumă la o simplă medie aritmetică, ci se ţine seama de o medie aritmetică ponderată. În cazul metaanalizei efectelor fixe, ponderarea variază pe o amplitudine mai mare, acordându-se o pondere mai mare studiilor cu mulţi subiecţi şi o pondere mai mică studiilor cu un număr redus de participanţi. În cazul metaanalizei efectelor aleatorii, ponderarea variază în mai mică măsură în funcţie de volumul eşantionului (deşi se menţine o asemenea diferenţiere), coeficientul de ponderare fiind dependent şi de discrepanţa mărimii efectului observată într-un studiu faţă de media studiilor analizate.

Borenstein *et al.* (2009) prezintă consecinţele acestor diferenţieri: (a) volumul eşantionului afectează în mai mare măsură ponderile acordate în metaanaliza efectelor fixe decât în cea a efectelor aleatorii; (b) valoarea dispersiei, cea a erorii standard de măsurare şi cea a intervalului de încredere pentru media efectelor ponderate vor fi întotdeauna mai mari în cazul metaanalizei efectelor aleatorii decât a efectelor fixe, pentru că la dispersia cauzată de eroarea de eşantionare (dispersia intrastudiu) se

adaugă dispersia datorată diferenţelor calitative dintre populaţiile investigate (dispersia interstudii; vezi caseta 7.8).

Caseta 7.8. *Calculul ponderii acordate unui studiu în cazul unei metaanalize a efectelor random după Borenstein* et al. *(2009) – cazul comparaţiilor dintre medii*

d, k	mărimea efectului (d Cohen) – vezi formula în caseta 4.7; numărul de studii incluse în analiză (k)
w, w_{random}	w – ponderea asociată mărimii efectului (metaanaliza fixă); w_{random} – ponderea asociată mărimii efectului în metaanaliza efectelor random
T^2, v_d, I^2	T^2 este dispersia interstudii; v_d este dispersia intrastudii (sau dispersia mărimii efectului pentru metaanaliza efectelor fixe; vezi formula în caseta 4.7); I^2 este un indicator al eterogenităţii efectelor
Q, C	indicatori ai eterogenităţii indicatorilor de mărime a efectelor

$$w_{random} = \frac{1}{v_{d_random}} \quad \text{unde} \quad v_{d_random} = v_d + T^2 \quad \text{iar} \quad T^2 = \frac{Q-(k-1)}{C}$$

$$I^2 = \left(\frac{Q-(k-1)}{Q}\right) \times 100\%$$

$$Q = \sum_{i=1}^{k} wd^2 - \frac{\left(\sum_{i=1}^{k} wd\right)^2}{\sum_{i=1}^{k} w} \qquad C = \sum w - \frac{\sum w^2}{\sum w}$$

Observaţie! Formulele se aplică şi în celelalte tipuri de metaanaliză, cu înlocuirea lui d şi v_d cu indicatorii corespunzători (de exemplu, z şi v_z pentru metaanaliza corelaţiilor). Dacă T^2 este negativ, i se atribuie valoarea 0. Formulele de calcul pentru v_d etc. sunt prezentate în caseta 7.7.

7.3.3. Calcularea principalilor indicatori metaanalitici

Discuţiile tehnice prezentate la primii doi paşi ai metaanalizei pot fi tratate expeditiv de consumatorii de literatură metaanalitică neinteresaţi de modul efectiv de calcul. În schimb, această secţiune este importantă

pentru a înţelege modul de interpretare a rezultatelor unei metaanalize. Iată o analogie între metaanaliză şi informaţiile din statistica descriptivă.

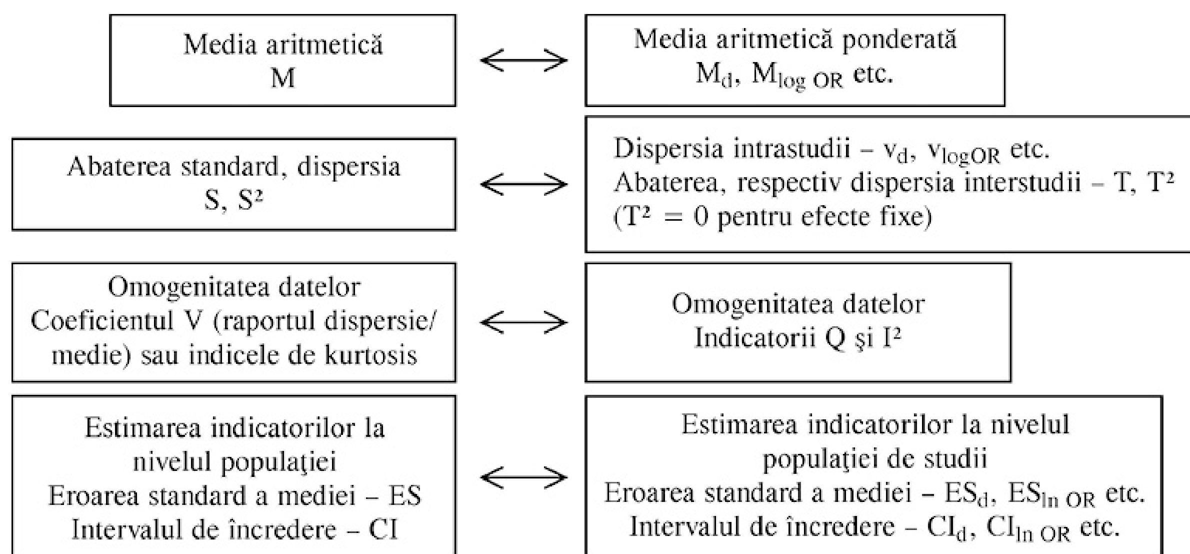


Figura 7.3. *Indicatorii descriptivi pentru un set de date (stânga) şi pentru metaanaliză (dreapta)*

Astfel, orice curs introductiv de statistică porneşte de la premisa că, pentru a descrie datele numerice obţinute la un eşantion de subiecţi, avem nevoie în primul rând de o medie şi o abatere standard care să descrie setul de date analizat. Suplimentar, se poate analiza forma distribuţiei pentru a observa, de exemplu, dacă datele colectate sunt foarte omogene (grupate în jurul mediei – distribuţie leptocurtică) sau sunt mai degrabă eterogene (foarte împrăştiate – distribuţie platicurtică). De asemenea, se poate calcula un interval de încredere pentru media obţinută, care ar indica valorile între care se situează media dacă din cadrul populaţiei vizate ar fi fost selectat alt set de participanţi (ţinând cont de erorile de eşantionare inerente). Aceeaşi logică este utilizată într-un demers metaanalitic pentru a găsi indicatorii descriptivi potriviţi pentru a descrie setul de studii analizate (prezentaţi schematic în figura de mai sus).

Pe baza acestor indicatori putem răspunde la următoarele întrebări (în paranteză fiind specificaţi indicii necesari):

– Efectul investigat este unul robust, semnificativ diferit de zero (media aritmetică ponderată, eroarea standard a mediei şi intervalul de încredere asociat mediei)?
– Efectul investigat este unul omogen – relativ similar – sau unul eterogen – variază mult de la un studiu la altul, fiind dependent de anumite caracteristici ale studiului (indicatorii Q şi I^2 ce estimează omogenitatea/eterogenitatea efectului)?

În cazul identificării unui efect eterogen se poate răspunde la alte două întrebări posibile: „Ce caracteristici ale studiilor contribuie la variaţii în mărimea efectului?”, respectiv „În termeni procentuali, cât la sută explică aceste caracteristici din modul de variere a mărimii efectului?”. Aceste întrebări suplimentare menite să explice motivele apariţiei eterogenităţii efectelor sunt discutate detaliat în secţiunea 7.3.4.

7.3.3.1. Interpretarea indicatorilor specifici metaanalizei

Valorile mediei aritmetice ponderate constituie unul dintre cei mai importanţi indicatori rezultaţi în urma demersului metaanalitic. Interpretarea acestui indicator statistic se poate face atât din perspectiva atât a statisticii descriptive, cât şi a celei inferenţiale.

Tabelul 7.2. *Semnificaţia valorilor mărimii efectului din perspectiva statisticii descriptive*

Tip de metaanaliză	Efect	Semnificaţie a valorilor (semne)	Valori scăzute	Valori medii	Valori ridicate
Comparaţii între medii	d, g	Rezultatele cu semnul plus indică diferenţe în direcţia aşteptată (de exemplu, $M_E – m_C > 0$), în timp	d = 0,20 g = 0,20 în modul	d = 0,50 g = 0,50 în modul	d = 0,80 g = 0,80 în modul

		ce rezultatele cu semnul minus indică diferenţe în direcţia contrară (de exemplu, $m_E – m_C < 0$). Valoarea zero indică absenţa diferenţelor dintre medii.			
Comparaţii între frecvenţe	OR, log OR	Rezultatele exprimate direct în raportul şanselor (OR) cu valoarea 1 indică absenţa diferenţelor dintre frecvenţe. Rezultatele supraunitare indică frecvenţe mai mari observate în grupul experimental (de exemplu, f succes E – f succes C > 0), iar rezultatele subunitare (între 0 şi 1) indică diferenţe de frecvenţe mai mici în grupul experimental. Rezultatele exprimate în log OR cu semnul plus indică diferenţe în direcţia aşteptată (de exemplu, f succes E – f succes C > 0). Rezultatele exprimate în log OR cu semnul minus indică diferenţe în direcţia contrară (de exemplu, f succes E – f succes C < 0). Valoarea zero indică absenţa diferenţelor dintre frecvenţe.	OR = 0,70 sau 1,45 log OR = 0,35 în modul	OR = 0,40 sau 2,50 log OR = 0,90 în modul	OR = 0,25 sau 4,00 log OR = 1,45 în modul
Corelaţii	r, z	Rezultatele cu semnul plus indică o asociere pozitivă între variabile, iar cele cu semnul minus semnifică o asociere negativă. Valoarea zero indică absenţa asocierii dintre cele două variabile.	r = 0,10 în modul	r = 0,25 în modul	r = 0,40 în modul

Din perspectiva statisticii descriptive, de interes este semnul şi magnitudinea mărimii medii a efectului (media aritmetică ponderată), detalii fiind prezentate în tabelul anterior. De exemplu, rezultatul unei

metaanalize corelaţionale ($r = 0,29$), în care media aritmetică ponderată a mărimii efectului din studiile analizate vizează relaţia dintre inteligenţă şi volumul (mărimea) creierului, semnifică existenţa unei corelaţii pozitive între variabilele studiate de intensitate medie (McDaniel, 2005). În mod similar, o valoare a raportului şanselor (OR) de 0,72 obţinută într-o metaanaliză ce compară frecvenţa mortalităţii pacienţilor care au beneficiat de consiliere psihologică faţă de cei care nu au beneficiat de acest serviciu în urma unei intervenţii chirurgicale pe cord indică un nivel de mortalitate mai scăzut (valori subunitare) în cazul pacienţilor care au beneficiat de consiliere psihologică. Un asemenea rezultat subliniază rolul benefic al consilierii psihologice în diminuarea implicării în comportamente riscante (fumat, consum de alcool, dietă neadecvată) la pacienţii care au suferit operaţii pe cord (Linden, Phillips şi Leclerc, 2007).

Din perspectiva statisticii inferenţiale, media aritmetică ponderată a mărimii efectului primeşte valenţe interpretative suplimentare, care ne ajută să răspundem la două întrebări: (a) „Este efectul unul robust/semnificativ statistic diferit de zero?”; (b) „Care este intervalul de încredere unde se regăseşte media obţinută în situaţia ipotetică în care am fi accesat întreaga populaţie de studii din domeniu?”.

Pentru a răspunde la prima întrebare este nevoie de raportul dintre media aritmetică ponderată şi eroarea standard a mediei, cea din urmă fiind rădăcina pătrată din suma dispersiei asociate studiilor (vezi caseta 7.7 în cazul utilizării calculului manual pentru metaanaliza efectelor fixe). Dacă acest raport (scor z) este mai mare sau egal cu 1,96 pentru un test bilateral, respectiv mai mare sau egal cu 1,65 pentru un test unilateral, putem spune că efectul găsit este semnificativ statistic diferit de zero, la un prag de semnificaţie $p < 0,05$. În acest caz, efectul este unul robust (de

exemplu, se poate afirma, într-adevăr, că există o asociere pozitivă între nivelul inteligenţei fluide şi volumul creierului)[8]. Mai mult, pe baza relaţiei dintre medie şi eroarea standard a mediei poate fi calculat intervalul de încredere al mediei ponderate, la diferite niveluri de probabilitate statistică.

Formula de calcul pentru aflarea intervalului de încredere are aceeaşi structură, indiferent de mărimea efectului măsurată. De exemplu, pentru cazul indicatorului d Cohen formula este:

Limita inferioară: $M_d – 1{,}96 \times SE_d$ Limita superioară: $M_d + 1{,}96 \times SE_d$

pentru 95% încredere

Limita inferioară: $M_d – 1{,}65 \times SE_d$ Limita superioară: $M_d + 1{,}65 \times SE_d$

pentru 90% încredere.

De exemplu, dacă media aritmetică ponderată într-o metaanaliză este $M_d = 0{,}65$, iar eroarea standard asociată acesteia este $SE_d = 0{,}17$, atunci intervalul de încredere (valorile între care se va regăsi cu o probabilitate de 95% media efectului la nivelul populaţiei de studii) va fi situat între 0,32 şi 0,98.

Desigur, cu cât acest interval de încredere este mai larg, cu atât este mai redusă utilitatea practică a efectului exprimat sub forma mediei aritmetice ponderate. Pentru a judeca mai obiectiv acest lucru, în interpretarea datelor se ţine seama de indicatorii de

omogenitate/eterogenitate. Valoarea mediei ponderate are o însemnătate practică mai ridicată (efectul fiind unul de încredere) dacă valorile obţinute indică mai degrabă o situaţie de omogenitate decât una de eterogenitate. Pentru a aprecia nivelul de omogenitate sau eterogenitate a efectelor utilizate în calculul mediei aritmetice ponderate se va ţine cont de doi indicatori suplimentari: Q, respectiv I^2.

Formula de calcul a lui Q a fost prezentată în caseta 7.8. Rezultatul obţinut este interpretat în contextul unei distribuţii χ^2 pentru k-1 grade de libertate, k reprezentând numărul de indicatori ai mărimii efectului incluşi în calculul mediei ponderate, cel mai adesea acesta fiind egal cu numărul de studii incluse în analiză. Dacă rezultatul este unul semnificativ statistic ($p < 0{,}05$) acest lucru este interpretat ca un semn al existenţei unei situaţii de eterogenitate, adică al unei variaţii destul de puternice în magnitudinea mărimii efectului de la un studiu la altul, în funcţie de anumite caracteristici ale acestora. În această situaţie, media aritmetică ponderată pentru toate studiile incluse în analiză îşi pierde din rolul sintetizator. În schimb, un rezultat Q nesemnificativ statistic ar putea fi interpretat ca un caz de omogenitate al efectelor analizate, aceste efecte variind în mai mică măsură de la un studiu la altul. Pentru a înţelege rolul omogenităţii sau eterogenităţii datelor, vom apela, prin analogie, la cazul unei medii aritmetice de 7. Aceasta poate fi obţinută în contextul a două note eterogene, 10 şi 4, situaţie în care, deşi media este una corectă din punct de vedere matematic, ea este mai puţin utilă pentru a estima rezultatele viitoare în comparaţie cu o medie de 7 obţinută din două note foarte omogene, precum ar fi chiar cazul a două note de 7. Din păcate, indicatorul Q este foarte sensibil la numărul de studii incluse în analiză. Dacă sunt incluse puţine studii (de exemplu, 10) puterea statistică de a obţine un Q semnificativ statistic este una foarte scăzută, ceea ce ar

însemna că vor putea exista situaţii de eterogenitate, deşi valoarea Q nu va fi una semnificativă statistic la un $p < 0,05$. Dacă sunt incluse multe studii (de exemplu, 100), puterea statistică de a obţine un Q semnificativ statistic este ridicată, ceea ce înseamnă că vor exista situaţii cu efecte omogene, semnalate ca fiind eterogene ($p < 0,05$). Deşi cea de-a doua situaţie discutată este mai puţin întâlnită, primul caz, cel al realizării metaanalizei pornind de la un număr relativ restrâns de studii, este mai frecvent întâlnit, ceea ce face dificilă interpretarea lui Q atunci când valoarea sa este nesemnificativă statistic. Avem de-a face într-adevăr cu un efect omogen sau rezultatul este nesemnificativ statistic din cauza puterii statistice scăzute asociate lui Q?

De aceea, Borenstein *et al.* (2009) recomandă ca, pe lângă interpretarea lui Q, să se ţină seama şi de interpretarea lui I^2 (vezi formula de calcul în caseta 7.8). Acesta ia valori între 0% şi 100% şi ar putea fi definit ca procentul din dispersia observată care indică diferenţe reale între valorile indicatorilor de mărime a efectului, nu doar variaţii aleatorii (*noise*). O altă semnificaţie a lui I^2 este aceea a unui indicator al inconsistenţei efectelor (a nivelului în care acestea variază destul de mult de la un studiu la altul). Avantajul lui I^2 asupra lui Q este faptul că I^2 nu este dependent de numărul de studii incluse în analiză, iar dezavantajul constă în faptul că I^2 nu se supune unui test de semnificaţie statistică pentru a aprecia dacă nivelul de eterogenitate este dovedit la un prag de semnificaţie statistică, fie el $p < 0,05$. Un alt dezavantaj important al lui I^2 este faptul că acesta reflectă doar procentul din dispersie care poate fi explicat (prin variaţii în caracteristicile studiului), însă nu ne spune nimic despre cât de eterogen este efectul găsit (informaţii pe care le conţine Q sau T^2; vezi caseta 7.8). În privinţa lui I^2, Higgins, Thompson, Deeks şi Altman (2003) propun următoarele valori ca reper pentru a estima măsura

în care eterogenitatea observată poate fi explicată de caracteristici ale studiilor incluse în analiză: 25% (nivel scăzut de variaţie sistematică a mărimii efectelor), 50% (mediu) şi 75% (ridicat).

Aşadar, din perspectiva interpretării dispersiei indicatorilor de mărime a efectului, din combinarea datelor obţinute pentru Q şi I^2 pot fi obţinute cel puţin patru situaţii ipotetice:

A. Q semnificativ statistic asociat cu un I^2 scăzut (deşi efectul obţinut este unul eterogen, variind destul de mult de la un studiu la altul, variaţiile observate se datorează mai degrabă erorii de eşantionare; prin urmare, ar fi neindicat să analizăm ce caracteristici ale studiilor contribuie la diferenţele observate în magnitudinea mărimii efectului, deoarece acestea nu contribuie prea mult la dispersia efectelor observate);

B. Q semnificativ statistic asociat cu un I^2 ridicat (efectul este unul eterogen, iar diferenţele observate de la un studiu la altul sunt explicate în mare măsură de variaţii în caracteristicile acestora, situaţie în care se cuvin a fi realizate analize suplimentare de tipul metaregresiei etc. (vezi secţiunea 7.3.4);

C. Q nesemnificativ statistic asociat cu un I^2 scăzut (efectul este unul omogen, variaţiile mici în magnitudinea mărimii efectului de la un studiu la altul fiind mai degrabă cauzate de erori de eşantionare; această interpretare ar trebui însă nuanţată în cazul unei analize făcute pe un număr redus de studii, unde se constată o putere statistică scăzută de a testa Q, situaţie în care amplitudinea intervalului de încredere pentru efectul mediu calculat ar putea suplini evaluarea gradului de eterogenitate a rezultatelor);

D. Q nesemnificativ statistic asociat cu un I^2 ridicat (efectul este omogen, dar chiar şi în aceste condiţii pot fi identificate anumite caracteristici ale studiilor responsabile pentru variaţii în magnitudinea efectului;

interpretarea lui Q poate fi nuanţată în termeni similari cu cei precizaţi la punctul C, în contextul unei puteri statistice reduse pentru analiza statistică a lui Q.

Pe scurt, Q secondat de T^2 sau de amplitudinea intervalului de încredere pentru media obţinută ne oferă informaţii despre cât de omogen sau eterogen este rezultatul mediu obţinut (media aritmetică ponderată), în timp ce I^2 ne indică măsura în care dispersia observată poate fi explicată (dispersie sistematică) sau nu (dispersie cauzată de erorile inerente în eşantionare). Dacă dispersia explicată reprezintă un procent însemnat (peste 50%), analizele secundare pot identifica sursele care explică variaţiile în mărimea efectului observat (relaţia dintre anumite caracteristici ale studiilor şi magnitudinea efectelor observate).

7.3.3.2. Ilustrarea interpretării rezultatelor unei metaanalize

Pentru a ilustra cele menţionate anterior vom face referire la cazul relaţiei dintre două trăsături de personalitate (neuroticismul, N, şi conştiinciozitatea, C) şi absenteismul cauzat de alte motive decât cele medicale sau decât cele determinate de anumite situaţii de urgenţă. În acest demers vom porni de la premisa ipotetică că au fost identificaţi doar 9 indicatori ai mărimii efectului pentru relaţia dintre N şi absenteism, respectiv doar 7 indicatori pentru relaţia dintre C şi absenteism (datele brute sunt prezentate în tabelul 7.3).

Fie în urma unui demers manual de calculare a indicatorilor necesari (casetele 7.7 şi 7.8), fie în urma utilizării unui software specializat,

precum este Comprehensive MetaAnalysis (Borenstein *et al.*, 2011), se obţin următoarele rezultatele prezentate în tabelul 7.4.

După cum se observă, numărul mic de studii incluse în metaanaliză ne obligă la utilizarea metaanalizei efectelor fixe, deşi din perspectivă teoretică o premisă realistă ar fi luat în calcul existenţa unei eterogenităţi în magnitudinea efectelor, deoarece este foarte probabil să existe efecte diferite într-un mediu şcolar (populaţie de elevi şi studenţi, organizaţii nonprofit) faţă de un mediu organizaţional (populaţie de adulţi, organizaţii orientate spre profit).

La nivelul rezultatelor obţinute privind relaţia dintre neuroticism şi absenteism se observă o corelaţie pozitivă robustă, de mică intensitate ($r = 0,158$, $p < 0,05$, deoarece valoarea asociată a lui z de 6,39 este semnificativă statistic) şi consistentă (omogenă; Q nesemnificativ statistic, iar distanţa dintre extremele intervalului de încredere este sub 0,10). De asemenea, se constată că un procent foarte mic din dispersia mărimii efectului în studii este cauzat de diferite caracteristici ale studiilor, cea mai mare parte (100% – 9,91% = 90,09%) fiind determinată de variaţii random (erori de eşantionare). În consecinţă, acest efect robust de mică intensitate este unul de încredere, indicând faptul că persoanele cu un nivel ridicat de neuroticism tind să acumuleze mai multe absenţe în mediul instituţional în care îşi desfăşoară activitatea, nefiind necesare alte analize complementare.

Tabelul 7.3. *Date brute şi prelucrări parţiale pentru exemplul oferit*

ID	%femei	M_age	Lot	N	r_{N_Ab}	r_{C_Ab}	z_{N_Ab}	z_{C_Ab}	w_N	w_C	$z \times w_N$	$z \times w_C$
1a	47	12,6	1	248	0,18	–0,16	0,182	–0,161	245	245	44,59	–39,44
1b	54	15,4	1	321	0,16	–0,13	0,161	–0,131	318	318	51,20	–41,66
1c	52	17,8	1	282	0,23	–0,16	0,234	–0,161	279	279	65,29	–44,92
2	70	36,2	2	54	–0,11	0,26	–0,110	0,266	51	51	–5,61	13,56

3	78	43,4	2	73	0,01	–0,24	0,010	–0,245	70	70	0,70	–17,15
4	14	42,6	2	181	0,19	–0,23	0,192	–0,234	178	178	34,16	–41,65
5	0	23,0	2	150	0,21	–0,28	0,213	–0,288	147	147	31,31	–42,33
6	83	42,3	2	210	0,08	–	0,080	–	207	–	16,57	–
7	100	28,7	2	113	0,16	–	0,161	–	110	–	17,71	–
				1.632				Total	1.605	1.288	255,92	–213,59

Legendă: ID – identificator studiu; %femei – procentul persoanelor de gen feminin incluse în studiu; M_age – vârsta medie a participanţilor la studii; Lot – tipul de participanţi incluşi (1 – elevi şi studenţi [şcoală]; 2 – angajaţi [organizaţii orientate spre profit]); N – număr de participanţi la studiu; r_{N_Ab} şi r_{C_Ab} – corelaţia dintre neuroticism şi absenteism, respectiv dintre conştiinciozitate şi absenteism; z_{N_Ab} şi z_{C_Ab} – valorile corelaţiilor convertite în z Fischer pentru neuroticism şi absenteism, respectiv pentru conştiinciozitate şi absenteism; w_N şi w_C – ponderile utilizate pentru neuroticism, respectiv conştiinciozitate, în calculul mediilor aritmetice ponderate; $z \times w_N$ şi $z \times w_C$ – produsul dintre ponderile fiecărui studiu şi valorile z din acel studiu (pentru neuroticism, respectiv conştiinciozitate). Suma produselor împărţită la suma ponderilor acordate oferă mediile aritmetice ponderate pentru analizele efectuate. De aceea, valorile mediei aritmetice ponderate pentru z Fisher sunt: pentru N ($\Sigma z \times w_N / \Sigma w_N$), respectiv pentru C ($\Sigma z \times w_C / \Sigma w_C$), iar valorile lui r se obţin după convertirea lui z în r (vezi caseta 7.7). Valorile obţinute sunt r = 0,158 pentru relaţia dintre N şi absenteism, respectiv r = –0,164 pentru relaţia dintre C şi absenteism).

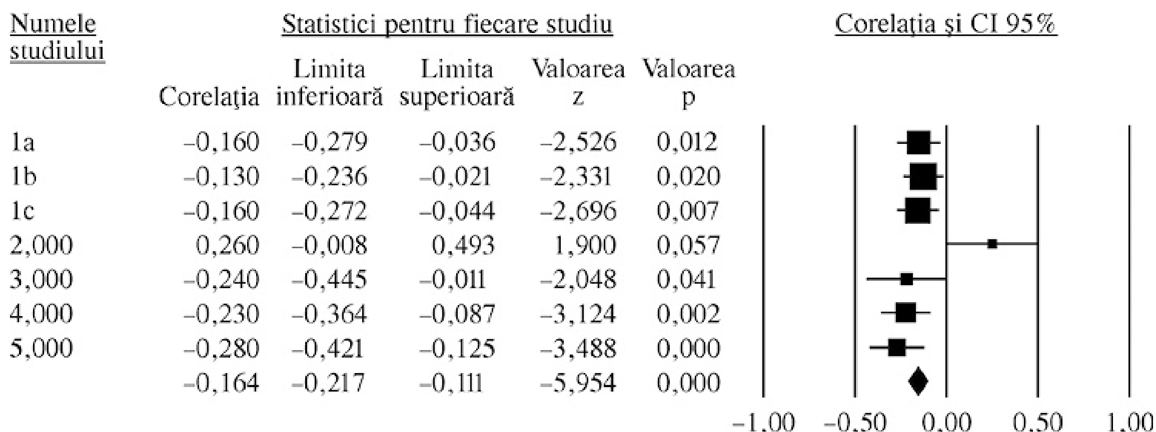
Tabelul 7.4. *Rezultatele metaanalizei corelaţiei dintre personalitate (N şi C) şi absenteism (metaanaliza efectelor fixe)*

Aspectul vizat	N	k	r	SD_r (T)	CI 95%	z	Q	I^2
Neuroticism – Absenteism	1632	9	0,158	0,025	[0,110; 0,206]	6,39*	8,88	9,91
Conştiinciozitate – Absenteism	1309	7	–0,164	0,084	[–0,217; –0,111]	5,95*	13,36*	55,11

Legendă: N – numărul total de participanţi; k – numărul de indicatori ai mărimii efectului incluşi în analiză (număr de studii independente analizate); r – media ponderată observată pentru mărimea efectului; SD_r sau T – abaterea standard asociată mediei ponderate; CI 95% – intervalul de încredere al mediei cu o probabilitate de 95%; z – testul de semnificaţie statistică (* pentru $p < 0,05$); Q – indicatorul eterogenităţii studiilor (* pentru $p < 0,05$); I^2 – procent din dispersia studiilor datorat altor factori decât erorii de eşantionare

În privinţa relaţiei dintre conştiinciozitate şi absenteism se observă că există o corelaţie negativă semnificativă statistic ($r = -0,164$, $p < 0,05$), însă efectul nu este unul consistent, robust (deoarece Q este semnificativ statistic), el variind într-o măsură moderată în funcţie de unele caracteristici ale studiilor analizate ($I^2 = 55,11\%$). În aceste condiţii, valoarea observată a mediei ponderate are puţină aplicabilitate practică. De aceea, se impune realizarea unor analize suplimentare pentru a identifica ce caracteristici ale studiilor au impact asupra modului în care variază efectul observat (joacă rolul unor variabile moderatoare). De exemplu, este posibil ca efectul eterogen cu privire la relaţia dintre conştiinciozitate şi absenteism să fie explicat de tipul de participanţi incluşi în studii (şcolari *vs* angajaţi), situaţie în care tipul participanţilor ar constitui o variabilă moderatoare. Într-un asemenea caz, analizele ar putea fi reluate separat pentru studiile efectuate în mediul educaţional, respectiv pentru studiile efectuate pe angajaţi, iar în urma acestui demers să constatăm efecte omogene la nivelul subgrupului de studii.

O modalitate intuitivă de a prezenta o parte însemnată dintre indicatorii descriptivi menţionaţi mai sus este de a reprezentarea fiecare studiu inclus în metaanaliză sub forma unui grafic care poartă denumirea, în limba engleză, de *forest plot*.



Numele studiului	Statistici pentru fiecare studiu				
	Corelaţia	Limita inferioară	Limita superioară	Valoarea z	Valoarea p
1a	–0,160	–0,279	–0,036	–2,526	0,012
1b	–0,130	–0,236	–0,021	–2,331	0,020
1c	–0,160	–0,272	–0,044	–2,696	0,007
2,000	0,260	–0,008	0,493	1,900	0,057
3,000	–0,240	–0,445	–0,011	–2,048	0,041
4,000	–0,230	–0,364	–0,087	–3,124	0,002
5,000	–0,280	–0,421	–0,125	–3,488	0,000
	–0,164	–0,217	–0,111	–5,954	0,000

Figura 7.4. *Descrierea grafică a studiilor incluse în metaanaliza relaţiei dintre conştiinciozitate şi absenteism*

După cum se observă din acest grafic (figura 7.4), în şase dintre cele şapte loturi analizate corelaţia obţinută între cele două variabile (conştiinciozitate şi absenteism) este una negativă. Mărimea pătratelor indică ponderea studiilor respective în calculul mediei ponderate a efectelor (studiile cu pătrate mai mari fiind cele în care au fost investigaţi mai mulţi participanţi şi în care dispersia rezultatelor a fost mai mică). Lungimea liniei se referă la intervalul de încredere asociat mărimii efectului din studiul respectiv, iar sub formă de romb este prezentată valoarea mediei ponderate a mărimii efectului. Software-uri specializate, precum Comprehensive MetaAnlysis (Borenstein *et al.*, 2011), obişnuiesc să asocieze datele descriptive prezentate cu o asemenea reprezentare grafică.

7.3.4. Indicatori metaanalitici suplimentari

În situaţia în care rezultatele obţinute nu sunt robuste sau în cazul în care efectelor robuste li se asociază o valoare I^2 ridicată, se obişnuieşte să se efectueze o serie de analize statistice suplimentare pentru a explica sursele variaţiei efectelor observate[9]. În acest sens se pot avea în vedere anumite caracteristici ale eşantioanelor incluse (de exemplu, caracteristici ale eşantioanelor utilizate: procent de femei din lotul analizat; vârsta medie a respondenţilor; tipul de respondenţi – şcolari *vs* angajaţi; vezi tabelul 7.3) sau caracteristici ale designului şi metodelor utilizate (studii transversale *vs* studii longitudinale; gruparea studiilor în funcţie de instrumentele psihometrice utilizate etc.). Aceste caracteristici ale studiilor sunt tratate ca variabile cu potenţial moderator. Pentru exemplul oferit vom analiza măsura în care eterogenitatea mărimii efectelor observate în cazul relaţiei

dintre conştiinciozitate şi absenteism poate fi explicată de diferenţe la nivelul caracteristicilor eşantioanelor (diferenţe între studii în ceea ce priveşte vârsta sau sexul participanţilor; diferenţe privind tipul respondenţilor: angajaţi *vs* şcolari). În cazul în care aceste variabile analizate (de exemplu, vârsta) sunt exprimate numeric, procedeul clasic de analiză se numeşte metaregresie. Prin intermediul acesteia putem decela măsura în care variaţia efectelor observate poate fi explicată de diferenţe în variabila predictor (diferenţe de vârstă). În cazul în care variabilele analizate sunt de tip categorial (de exemplu, şcolari *vs* angajaţi), se apelează la o analiză a diferenţelor dintre subgrupuri de studii, într-un mod similar, dar nu identic, cu cel întâlnit în cazul tehnicilor ANOVA sau al testelor t pentru grupuri independente.

Înainte de a prezenta principalele particularităţi implicate într-o metaregresie sau într-o analiză de subgrupuri, se cuvine să precizăm o serie de argumente oferite de Borenstein *et al.* (2009) în favoarea recomandării de a nu apela la tehnicile de lucru consacrate din statistica tradiţională, de tipul regresiei liniare sau al tehnicilor ANOVA, ci de a apela la aceste tehnici specifice, denumite metaregresii, respectiv la analiza subgrupurilor.

În primul rând, tehnicile statistice clasice, utilizate de mulţi autori ca soluţie practică de a analiza complementar relaţia dintre caracteristicile studiilor şi magnitudinea efectelor, se realizează în condiţiile acordării unei ponderi egale fiecărui indicator observat al mărimii efectului, fără a se ţine astfel seama de diferenţele ca pondere (mărime) ale fiecărui studiu. Apoi, chiar dacă acest neajuns ar fi corectat prin apelul la sintaxe, astfel încât analizele să ia în calcul ponderi diferite pentru studiile incluse, analizele respective nu ar diferenţia între analizele bazate pe efecte fixe şi cele bazate pe efecte random sau mixte. De asemenea, o altă diferenţă ar

viza modul de stabilire a gradelor de libertate, care se face diferit în tehnicile tradiţionale comparativ cu metaregresia sau analiza de subgrupuri, iar acest aspect poate uneori influenţa rezultatul final obţinut, prin raportarea la praguri de semnificaţie diferite.

7.3.4.1. Metaregresiile

Atunci când vrem să identificăm în ce măsură anumite caracteristici ale studiilor incluse, exprimate numeric (de exemplu, vârsta medie a participanţilor), sunt asociate cu variaţii în magnitudinea mărimii efectului găsit, se apelează la metaregresii. Acestea pot ţine cont în calculul lor de tipul de efecte urmărite (fixe *vs* random). Dintre cele două, cel mai plauzibil model de lucru este cel al lucrului cu efecte random, deoarece în acest caz se porneşte de la premisa că variabila moderatoare analizată explică doar o parte a variaţiilor observate la nivelul mărimii efectului, restul fiind pus pe seama altor predictori (variabile moderatoare) neanalizaţi. Cu alte cuvinte, în modelul random din metaregresie se porneşte de la premisa că intensitatea mărimii efectului variază nu doar din cauza unor erori de eşantionare (dispersia erorii), ci şi din cauza mai multor caracteristici ale studiilor analizate, printre care şi variabila vizată în metaregresia respectivă. În schimb, metaanaliza efectelor fixe porneşte de la premisa că variaţiile observate la nivelul mărimii efectului se datorează în întregime doar erorilor de eşantionare şi diferenţelor dintre studii existente exclusiv la nivelul caracteristicii incluse în metaregresie.

Totuşi, ca şi în alte cazuri menţionate anterior, rămâne recomandarea de a aplica modelul efectelor fixe în cazul unei metaregresii efectuate pe puţine studii, deşi teoretic ar fi mai potrivit un model al efectelor random.

Acest lucru derivă din specificul calculului efectelor random care include un parametru bazat pe dispersia dintre studii (T^2), care este un indicator uşor influenţabil (poate varia semnificativ) atunci când sunt incluse un număr redus de studii în analiză.

Tabelul 7.5. *Rezultatele metaregresiei pentru a surprinde relaţia dintre anumite caracteristici ale eşantioanelor utilizate şi magnitudinea corelaţiei dintre personalitate (C) şi absenteism (perspectiva metaanalizei efectelor random)*

Aspectul vizat	Valoare	Eroare standard	z	p	Q_{model}	p
Intercept procentaj femei	–0,2818	0,0611	–4,61*	0,000		
Procentaj femei	0,0027	0,0012	2,13*	0,033	4,538*	0,033
Intercept vârsta participanţilor	–0,1428	0,0627	–2,27*	0,022		
Vârsta participanţilor	–0,0010	0,0025	–0,40	0,681	0,168	0,681

* Situaţiile în care valoarea lui z este semnificativă statistic la un $p < 0,05$

Procedurile de calcul sunt destul de complicate, utilizatorii interesaţi putând consulta Borenstein *et al.* (2009) pentru detalii despre algoritmii de calcul. În cele ce urmează voi prezenta doar interpretarea rezultatelor obţinute printr-o asemenea procedură – de exemplu, în cazul utilizării unui software precum Comprehensive MetaAnalysis (Borenstein *et al.*, 2011) – pentru cei doi predictori numerici care se pretează pentru analiza de tip metaregresie: procentul de femei din studiu (relaţia dintre gen şi magnitudinea efectelor observate), respectiv vârsta medie a participanţilor din studii (relaţia dintre vârstă şi magnitudinea efectelor studiate). Interpretarea se face într-un mod similar cu cel întâlnit în cazul regresiei liniare simple, contând semnul valorii calculate şi dacă această valoare este sau nu una semnificativă statistic, în baza testului z.

Rezultatele de mai sus indică existenţa unei asocieri între procentajul de femei incluse în studiu şi intensitatea legăturii dintre conştiinciozitate

şi absenteism. Se observă că există tendinţa de a obţine valori mai ridicate ale corelaţiilor dintre conştiinciozitate şi absenteism în studii în care procentajul de femei este mai ridicat (vezi semnul plus al valorii asociate procentajului de femei = 0,0027). Acestei valori îi este asociat un test z semnificativ statistic ($p < 0,05$). În schimb, se observă că vârsta medie a participanţilor nu este asociată cu variaţii în magnitudinea relaţiei dintre conştiinciozitate şi absenteism (valorii –0,0010 îi corespunde o valoare z nesemnificativă statistic). În situaţia ipotetică în care acestei valori negative i-ar fi corespuns un z semnificativ statistic ($p < 0,05$), semnul minus ar fi fost interpretat în sensul unei corelaţii negative între vârstă şi intensitatea relaţiei dintre conştiinciozitate şi absenteism.

La aceeaşi interpretare se poate ajunge analizând un alt indicator oferit de Comprehensive MetaAnalysis (Borenstein *et al.*, 2011), precum este Q_{model} (dispersia determinată de diferenţele dintre studii). Dacă rezultatul este unul semnificativ statistic, înseamnă că acea caracteristică a studiului este relevantă pentru a prezice relaţia dintre conştiinciozitate şi absenteism, fără a preciza însă direcţia acelei relaţii (dacă conduce la creşterea sau descreşterea legăturii). Direcţia este dată de semnul asociat variabilei-predictor în cadrul metaregresiei, aşa cum s-a exemplificat anterior.

7.3.4.2. Analiza subgrupurilor

Atunci când caracteristicile studiilor analizate nu sunt exprimate numeric pentru a permite derularea de metaregresii, ci sub forma unor variabile nominale, se apelează la analiza diferenţelor la nivel de subgrupuri de studii. Un astfel de caz este cel în care studiile ce vizau relaţia dintre conştiinciozitate şi absenteism sunt grupate în două categorii: studii

realizate în mediul educaţional, pe elevi sau studenţi, *vs* studii realizate în întreprinderi, pe angajaţi.

Spre deosebire de metaregresie, care presupune lucrul cu efecte fixe sau efecte random, în cazul analizei de subgrupuri sunt posibile trei opţiuni de lucru, deoarece se analizează două niveluri ierarhice diferite (analiza studiilor din acelaşi subgrup – de exemplu, analiza făcută la nivelul studiilor bazate pe angajaţi, respectiv analiza care vizează diferenţele la nivel de subgrupuri, respectiv analiza care vizează surprinderea unor diferenţe între grupuri, în acest caz şcolari *vs* angajaţi). De aceea, cele trei variante de lucru sunt modelul efectelor fixe, modelul efectelor random şi modelul efectelor mixte. Această ultimă posibilitate apare, de exemplu, atunci când cercetătorul porneşte de la premisa utilizării efectelor random în interiorul subgrupurilor, dar utilizează modelul efectelor fixe pentru a analiza diferenţele dintre subgrupuri.

Ca şi în cazul metaregresiilor, rezultatele pot fi interpretate pornind de la valorile lui z sau ale lui Q (dispersie a diferenţelor dintre subgrupe de studii). O altă variantă simplificată de lucru ar fi calculul lui Q pentru nivelul de eterogenitate (vezi caseta 7.8) luând în calcul media grupurilor comparate. În acest context se calculează mai întâi indicii pentru fiecare subgrup în parte, derulându-se în paralel mai multe metaanalize în funcţie de numărul de subgrupe vizate. În exemplul oferit vom calcula parametrii specifici pentru studiile ce au avut o populaţie şcolară, respectiv pentru studiile realizate pe angajaţi. Rezultatele din cele două subgrupe sunt apoi analizate, ca şi cum ar fi provenit din două studii, prin prisma gradului de omogenitate sau eterogenitate a lor. Dacă efectul Q este unul eterogen (semnificativ statistic), înseamnă că acea caracteristică a studiului evaluată (în cazul dat, tipul de participanţi: şcolari *vs* angajaţi) moderează magnitudinea efectului găsit.

Apelând la metaanaliza efectelor fixe la nivel de subgrupuri (fiind doar 7 studii incluse în analiză, dintre care 3 vizează populaţie şcolară, iar 4 se referă la angajaţi), rezultatele obţinute cu ajutorul software-ului Comprehensive MetaAnalysis (Borenstein *et al.*, 2011) sunt prezentate în tabelul 7.6.

Tabelul 7.6. *Rezultatele analizei de subgrupuri privind impactul tipului de participanţi analizaţi (şcolari* vs *angajaţi) asupra magnitudinii efectului ce vizează corelaţia dintre conştiinciozitate (C) şi absenteism (perspectiva metaanalizei efectelor fixe)*

Relaţia C – absenteism	k	r	SD_r (T)	CI 95%	z	Q	df	I²
A. Şcolari	3	–0,149	0,000	[–0,214; –0,082]	4,34*	0,18	2	0,00%
B. Angajaţi	4	–0,194	0,176	[–0,281; –0,103]	4,14*	12,55*	3	76,10%
Total	7	–0,164	0,084	[–0,217; –0,111]	5,95*	13,36*	6	55,11%

Datele din tabel nu indică direct dacă între coeficienţii asociaţi celor două subgrupuri există diferenţe semnificative statistic. În schimb, datele din acest tabel indică faptul că, atât la nivel global, cât şi separat pe loturi de şcolari, respectiv angajaţi, efectul este unul semnificativ statistic, cu semn negativ (ceea ce indică o relaţie negativă între nivelul de conştiinciozitate şi absenteism). Pentru a afla dacă diferenţele sesizate la nivelul mediei ponderate a mărimii efectului la şcolari (r = –0,149), respectiv angajaţi (r = –0,194) sunt suficient de mari încât să indice diferenţe semnificative statistic, se poate face un calcul suplimentar pe baza datelor oferite în tabelul anterior.

În acest sens, se calculează un Q şi un df pentru diferenţa dintre grupe, ţinând seama de diferenţa dintre valorile obţinute pentru toate cele 7 studii analizate şi suma valorilor obţinute în analizele efectuate pentru cele două categorii de participanţi (şcolari şi angajaţi).

$$Q_{between} = Q_{total} - (Q_A + Q_B) = 13{,}36 - (0{,}18 + 12{,}55) = 0{,}63$$

$$df_{between} = df_{total} - (df_A + df_B) = 6 - (2 + 3) = 1$$

Rezultatul obţinut în cazul de faţă ($Q = 0,63$ la un grad de libertate) este interpretat pe baza unei distribuţii hi pătrat, rezultatul fiind unul nesemnificativ statistic ($p > 0,05$), fapt pentru care se poate spune că tipul de participanţi incluşi în studiu nu afectează magnitudinea corelaţiei dintre conştiinciozitate şi absenteism.

7.3.4.3. Impactul procentual al variabilelor moderatoare

Indiferent dacă se apelează la metaregresii sau la analiza diferenţelor dintre subgrupuri, atunci când se utilizează un model al efectelor random (nu este şi cazul exemplului lucrat) se poate oferi un răspuns suplimentar. Mai concret, se poate aprecia procentual impactul unei variabile moderatoare analizate luând în calcul totalul variaţiilor observate în magnitudinea efectelor ce pot fi puse pe seama diferitelor caracteristici ale studiilor (procent din I^2). Acest indicator poate fi interpretat similar cu indicatorul R^2 întâlnit în cazul regresiilor liniare, luând valori între 0 şi 100%, cu precizarea că el explică procentual doar evoluţia dispersiei determinate de diferite caracteristici ale studiilor (dispersia intergrup), nu şi dispersia determinată de erorile de eşantionare (dispersia intragrup). Aşa cum am amintit, un astfel de indicator are sens doar în cazul unei metaanalize a efectelor random, deoarece în cazul metaanalizei efectelor fixe se porneşte de la asumpţia că diferenţele observate între valorile mărimilor efectului sunt cauzate exclusiv de erori de eşantionare.

7.3.5. Verificarea robusteţii rezultatelor obţinute

După discutarea principalelor rezultate obţinute (nivelul şi semnificaţia mediei ponderate, nivelul de eterogenitate a datelor) şi a surselor (caracteristici ale studiilor) care prezic variaţii în magnitudinea efectului, sunt necesare câteva analize suplimentare care pot fi împărţite în două mari categorii: (a) analiza sensibilităţii rezultatelor; (b) analize ale gradului de distorsiune a rezultatelor pentru a vedea în ce măsură rezultatele sunt afectate de studiile relevante neidentificate (nepublicate).

7.3.5.1. Analiza sensibilităţii rezultatelor

Acest tip de analiză urmăreşte compararea rezultatelor obţinute în metaanaliză cu cele obţinute în situaţia ipotetică în care am fi apelat la alte restricţii de selectare a studiilor. De exemplu, o metaanaliză privind efectul terapiei cognitiv-comportamentale asupra reducerii simptomatologiei depresive la adolescenţi ar putea fi realizată incluzând atât studiile de tip experimental (cu grup de control), cât şi pe cele de tip cvasiexperimental (cu grup de control). O analiză a sensibilităţii ar urmări robusteţea rezultatelor în contextul în care analiza cuprinde ambele categorii de studii sau doar pe cele experimentale. În exemplul nostru cu privire la relaţia dintre conştiinciozitate şi absenteism, să presupunem că trăsătura de personalitate a fost evaluată în cinci dintre cazuri cu proba NEO-FFI (Costa şi McRae, 1992), iar în alte două cazuri cu un alt instrument valid pentru evaluarea conştiinciozităţii. O analiză de sensibilitate ar putea viza în ce măsură rezultatele metaanalizei converg în aceeaşi direcţie în cazul analizării tuturor celor şapte studii faţă de aceea în care am cuprinde doar cele cinci studii în care s-a utilizat NEO-FFI ca instrument de evaluare a personalităţii. Pe scurt, prin analiza de sensibilitate se urmăreşte în primul rând să se evidenţieze impactul pe

care diferite decizii de selectare (criterii de includere) a studiilor sau anumite decizii de calcul (efecte fixe sau random) îl au asupra rezultatului metaanalizei. Desigur, ideal ar fi ca efectele obţinute să fie similare (robuste) indiferent de variaţiile luate în calcul printre criteriile de includere a studiilor sau printre metodele de calcul utilizate. Analizele de moderare menţionate în secţiunea 7.2.4 pot fi utilizate pentru a verifica robusteţea efectelor, fiind de dorit să nu existe un efect moderator semnificativ statistic.

7.3.5.2. Analize ale gradului de distorsiune a rezultatelor în urma selecţiei studiilor

Oricât de mult s-ar strădui un cercetător să găsească toate studiile relevante pe o anumită tematică, există o probabilitate foarte mare ca unele studii să nu fie incluse, şi pentru simplul fapt că o parte dintre studiile realizate nu ajung să fie publicate sau măcar comunicate. În acest context, se pune problema măsurii în care rezultatele rămase în „sertar" afectează rezultatele metaanalizei. În acest scop au fost identificate o serie de proceduri menite să indice: (a) nivelul riscului de a întâlni un caz de distorsiune a rezultatelor ca urmare a eşecului de a include toate studiile eligibile; (b) nivelul robusteţii rezultatelor obţinute, în ciuda existenţei unei distorsiuni în rezultatele metaanalizei; (c) impactul distorsiunii asupra rezultatelor.

Pentru primul nivel al discuţiei, cel al identificării riscului de a avea rezultate distorsionate, modalitatea tipică utilizată este norul de puncte de tipul pâlniei întoarse (simetrice; *funnel plot*). În mod normal, dacă există un risc ridicat de distorsiune, atunci norul de puncte nu este unul simetric, ci există tendinţa de a observa mai multe studii cu o mărime a efectului

scăzută în cazul studiilor efectuate pe eşantioane mari de persoane (ce induc valori scăzute ale erorii standard de măsurare) decât în cazul studiilor efectuate pe eşantioane reduse ca volum de persoane. Acest lucru se întâmplă din cauza tendinţei multor publicaţii ştiinţifice de a accepta cu precădere articole în care se obţin rezultate semnificative statistic, asociate cu o probabilitate scăzută de a obţine un rezultat semnificativ statistic în contextul unui efect scăzut şi al unui eşantion redus de participanţi. De aceea, premisa de la care se porneşte este aceea că studiile rămase nepublicate pot distorsiona rezultatul metaanalizei.

În cazul în care nu sunt semne de distorsiune, graficul ar trebui să fie unul simetric, cu efecte mai ridicate sau mai scăzute obţinute indiferent de volumul eşantioanelor incluse în studiu, chiar dacă în cazul studiilor efectuate pe un număr mare de respondenţi variaţia efectelor tinde să scadă, deoarece valorile respective se apropie în mai mare măsură de media teoretică a efectului de la nivelul populaţiei. Un asemenea grafic se realizează în mod convenţional prin exprimarea valorii mărimilor efectelor pe axa orizontală şi a volumului eşantionului (sau a erorii standard de măsurare) pe axa verticală.

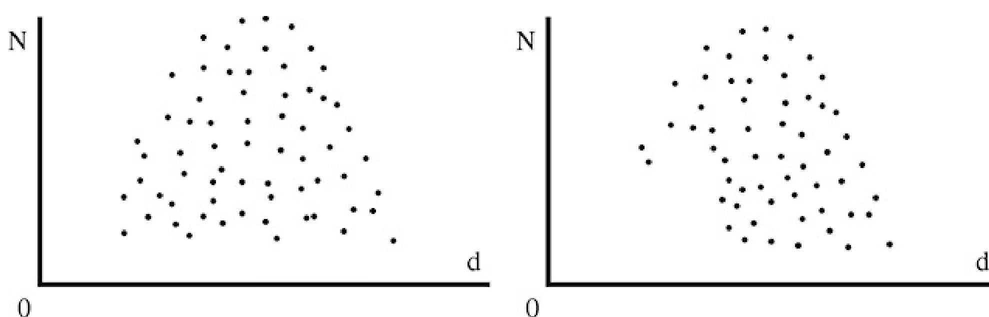


Figura 7.5. *Situaţie indicativă pentru absenţa distorsiunii de publicare (simetric; în stânga), respectiv pentru prezenţa distorsiunii de publicare (asimetric; în dreapta)*

Totuşi, deşi metoda este intuitivă, ea are câteva neajunsuri ce merită semnalate: (a) pentru a fi aplicată este nevoie de un număr destul de

ridicat de studii identificate (de exemplu, cel puţin 20), deoarece este dificil de apreciat gradul de simetrie/asimetrie atunci când numărul de studii este unul foarte redus; (b) asimetria/simetria este analizată subiectiv, la nivelul impresiei (percepţiei) cercetătorului, nefiind susţinută de indicatori concreţi; (c) în sfârşit, chiar dacă se constată asimetria datelor, nu se poate determina calitatea studiilor neincluse în analiză, fiind posibil ca acele efecte mici să fie cauzate de anumite probleme metodologice (variabile străine necontrolate), fapt care ar îndreptăţi excluderea studiilor respective din analiză. Într-un asemenea context nu ar fi corect să vorbim despre distorsiunea rezultatelor din metaanaliză, ci despre o decizie corectă de excludere a acestora.

În privinţa celei de-a doua limitări sesizate, cea legată de impresia subiectivă, unele programe specializate, precum Comprehensive MetaAnalysis, oferă indicatori suplimentari, cum ar fi coeficientul de corelaţie a rangurilor al lui Begg şi Mazumdar (1994) sau interceptul lui Egger *et al.* (1997). Găsirea unui rezultat semnificativ în urma calculării acestor indicatori obiectivi ar indica prezenţa unei distorsiunii de publicare, însă trebuie menţionat faptul că pragurile de semnificaţie asociate acestor indicatori sunt adecvate în contextul unui număr suficient de mare de studii luate în calcul. În caz contrar, pentru analize efectuate pe un număr redus de studii există riscul obţinerii unor rezultate nesemnificative statistic, din cauza unei puteri statistice reduse, ceea ce nu ar însemna neapărat absenţa distorsiunii determinate de neincluderea unor studii.

Prin urmare, pentru a aprecia corect riscul prezenţei unei distorsiuni în rezultatele metaanalizei, toate metodele menţionate mai sus funcţionează eficient doar în contextul existenţei unui număr destul de mare de studii incluse (cel puţin 20). Pentru alte situaţii, cum este cazul relaţiei dintre

conştiinciozitate şi absenteism în care au fost incluse doar şapte studii, aceşti indicatori sunt mai puţin relevanţi, din cauza puterii statistice reduse de a decela distorsiunile existente. Totuşi, în contextul obţinerii de indicatori semnificativi statistic şi a unui grafic de tip pâlnie evident asimetric, putem miza pe existenţa unei distorsiuni a rezultatelor inclusiv în cazul unui număr redus de studii incluse.

O întrebare adiacentă legată de problema distorsionării rezultatelor, mai ales în contextul în care se sesizează o asemenea tendinţă, este aceea a identificării numărului de studii de siguranţă (*fail-safe N*). În varianta sa mai cunoscută (clasică) propusă de Rosenthal (*apud* Borenstein *et al.*, 2009; vezi Radu *et al.*, 1993 pentru formula de calcul), se poate identifica numărul de studii care lipsesc având un efect egal cu zero (de exemplu, $r = 0{,}00$) necesar pentru a schimba media mărimii efectului dintr-un rezultat semnificativ statistic într-unul nesemnificativ statistic. Utilizând Comprehensive MetaAnalysis (Borenstein *et al.*, 2011) pentru a obţine acest indicator în exemplul dat, vom observa că ar fi nevoie de un număr de 47 de studii cu efect nul pentru a ajunge la concluzia că nu există o legătură între conştiinciozitate şi absenteism, respectiv de un număr de 66 de studii cu efect nul pentru a ajunge la concluzia că între nivelul de neuroticism şi absenteism nu există o legătură semnificativă.

O alternativă de calcul realistă şi mai adaptată cerinţelor practicienilor este cea propusă de Orwin. Acesta permite estimarea numărului de studii neincluse în metaanaliză care ar schimba valoarea mediei ponderate a mărimii efectului până la un anumit nivel minim acceptat (de exemplu, în cazul corelaţiilor un asemenea nivel ar putea fi $r = 0{,}10$, specific unui efect scăzut, cu utilitate practică redusă). Cu alte cuvinte, metoda propusă de Orwin identifică numărul de studii necesare pentru a transforma un efect observat în unul fără relevanţă practică deosebită, chiar dacă efectul

luat ca reper nu este nul. De asemenea, acest indicator îi oferă cercetătorului flexibilitatea de a specifica pentru studiile neincluse, pe lângă valoarea zero (absenţa totală a efectului), şi alte valori medii diferite de zero (de exemplu, un r mediu de 0,05 pentru studiile neincluse). Astfel, în contextul dat, ne-am putea întreba care este numărul necesar de studii neincluse pentru ca media ponderată a mărimii efectului să ajungă la un nivel minim acceptat de 0,10 (pentru relaţia dintre neuroticism şi absenteism), respectiv de –0,10 (pentru relaţia dintre conştiinciozitate şi absenteism), în contextul în care media efectelor din studiile neincluse ar fi nulă (r = 0,00). Utilizând un program specializat, precum Comprehensive MetaAnalysis (Borenstein *et al.*, 2011), am obţine următoarele valori: doar 5 studii neincluse cu efecte nule ar fi necesare pentru a reduce mărimea efectului la o valoare minim acceptabilă de –0,10 pentru relaţia dintre conştiinciozitate şi absenteism, respectiv doar 6 studii neincluse ar fi necesare pentru a reduce mărimea efectului privind relaţia dintre neuroticism şi absenteism la un nivel minim ca valoare practică (r = 0,10). Desigur, aceste valori sunt valabile în contextul în care am specificat că studiile neincluse ar avea o mărime a efectului egală cu zero. Interpretarea acestor rezultate se face contextual. De pildă, 5-6 studii neincluse semnifică un risc mare de distorsiune în date dacă cercetătorul reuşeşte să identifice 100 de studii eligibile, fiind posibil ca un procent de 5-6% din studii să scape analizei. Aceeaşi valoare de 5-6 studii neincluse înseamnă însă un risc mai scăzut de distorsiune în cazul în care au fost identificate mai puţin de 10 studii eligibile, deşi au fost respectaţi paşii de evaluare sistematică a literaturii de specialitate. Putem afirma acest lucru pentru că este puţin probabil să scape analizei un procent de 50-60% din totalul studiilor relevante, dacă evaluarea sistematică a fost realizată corect.

În sfârşit, o altă metodă de lucru în cazul identificării unui risc crescut de distorsiune a datelor este metoda tăierii şi adăugirii (*trim and fill*) propusă de Duval şi Tweedie (2000). Această procedură iterativă porneşte de la graficul de tip pâlnie întoarsă, încercând să adauge sau să elimine studii astfel încât să se obţină un grafic simetric. Desigur, o asemenea analiză este implementată doar dacă se constată o distribuţie asimetrică a efectelor. Din acest punct de vedere, demersul se apropie de ideea analizei sensibilităţii, răspunzându-se la întrebarea: „Care ar fi media ponderată a mărimii efectului dacă nu ar exista distorsiunea de includere selectivă a studiilor?”. Cu ajutorul programelor specializate, precum Comprehensive MetaAnalysis (Borenstein *et al.*, 2011), se pot compara direct indicatorii importanţi (medie, abatere standard, interval de încredere, eterogenitate), atât pentru cazul analizei rulate pe studiile identificate iniţial, cât şi pentru cazul analizei realizate pe un set modificat de studii (obţinut prin adăugarea/eliminarea unor efecte), astfel încât să se reflecte o distribuţie simetrică, ceea ce semnifică absenţa distorsiunii induse de studiile neincluse în analiză.

Concluzii

Apelul la evaluarea sistematică a unei tematici prin intermediul tehnicii metaanalitice constituie principala modalitate de sistematizare a rezultatelor dintr-un domeniu de interes. Metaanaliza nu are însă doar rolul unui demers sumativ, ci poate deschide noi direcţii de acţiune (ridică noi întrebări de investigat).

Calitatea rezultatelor unei metaanalize este dependentă de măsura în care cercetătorul a inclus toate studiile relevante pentru tema de

investigat, cât şi de criteriile de selecţie utilizate (de exemplu, metaanalizele care vizează identificarea unei relaţii de tip cauză-efect includ doar studii experimentale). Orice evaluare sistematică ar trebui să facă apel la cel puţin două baze de date independente şi să aplice strategii alternative de identificare a unor studii relevante.

În privinţa metaanalizei, deşi există o multitudine de indicatori statistici disponibili, cel mai adesea un demers metaanalitic complet va face referire la: (a) prezentarea mediei efectului, a abaterii standard şi/sau a intervalului de încredere pentru aceasta, precum şi a gradului de semnificaţie a acestui rezultat (semnificativ sau nesemnificativ statistic); (b) analiza nivelului de omogenitate/eterogenitate a efectului găsit (Q) şi a măsurii în care aceste variaţii pot fi explicate de caracteristici ale studiilor (I^2); (c) în cazul sesizării unor efecte eterogene ce pot fi explicate de caracteristici diferite ale studiilor incluse, metaanaliza se completează cu metaregresii şi analize de subgrupuri pentru a identifica rolul/impactul fiecărei caracteristici relevante în dinamica mărimii efectului analizat; (d) suplimentar, ar fi necesar să se facă referiri la analiza sensibilităţii şi a distorsionării rezultatelor, pentru a aprecia nivelul de robusteţe al rezultatelor exprimate.

În ceea ce priveşte metodele de lucru, cel mai potrivit model este cel al efectelor random, deoarece este puţin probabil ca efectele observate de-a lungul studiilor să varieze doar din cauza unor erori de eşantionare. În plus, acest model are avantajul de a extinde concluziile, dincolo de sfera studiilor analizate şi se pretează mai bine atunci când sunt cuprinse mai multe studii (de exemplu, cel puţin 15-20), lucru mai dificil de realizat în unele cazuri, mai ales dacă se doreşte realizarea unei serii de analize la nivel de subgrupuri.

Lecturi suplimentare

Michael Borenstein, Larry V. Hedges, Julian P.T. Higgins, Hannah R. Rothstein (2009). *Introduction to MetaAnalysis.* Chichester, UK: Wiley.

Această carte este fundamentală pentru cei care doresc să înţeleagă în profunzime aspectele prezentate pe parcursul acestui capitol, ea devenind foarte curând principala lucrare de referinţă în domeniul metaanalizei. Autorii cărţii sunt cei care au introdus şi programul adiacent menţionat în acest capitol, Comprehensive MetaAnalysis, acesta fiind disponibil gratuit, într-o variantă demonstrativă cu termen de valabilitate limitată.

Julian P.T. Higgins, Sally E. Green (2011). *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 (updated March 2011).* Disponibil la http://www.cochrane-handbook.org.

Referinţa este utilă îndeosebi celor din zona psihologiei clinice şi a psihoterapiei, fiind o lucrare necesară pentru a înţelege pas cu pas ce înseamnă şi cum se realizează efectiv o evaluare sistematică.

Mark W. Lipsey, David B. Wilson (2001). *Practical MetaAnalysis.* Thousand Oaks, SUA: Sage Publications.

Recomand lucrarea îndeosebi celor care au o formare umanistă şi care au întâmpinat dificultăţi în a înţelege concepte din zona metodologiei cercetării şi a statisticii sociale. Cartea este scrisă într-un limbaj accesibil, oferind o imagine de ansamblu atât asupra evaluării sistematice, cât şi asupra demersului metaanalitic propriu-zis.

Laurenţiu P. Maricuţoiu (2008). Introducere în metaanaliză. Metaanaliza corelaţiilor. *Psihologia Resurselor Umane*, 6, pp. 112-127.

Articolul detaliază fundamentele metaanalizei şi prezintă punctual paşii necesari pentru derularea unei metaanalize bazate pe studii corelaţionale, după modelul propus de Hunter şi Schmidt (1994/2001), un model de lucru răspândit mai ales în domeniul psihologiei organizaţionale, o zonă de cercetare în care predomină studiile de tip non-experimental.

Bibliografie

APA (2011). *PsycINFO – American Psychological Association*. Găsit la adresa: http://www.apa.org/ pubs/databases/psycinfo/index.aspx

Becker, L.A. (1999). *Effect size calculators*. Găsit la adresa http://www.uccs.edu/~faculty/lbecker/

Begg, C.B., Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, pp. 1088-1101.

Borenstein, M.. Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. (2009). *Introduction to MetaAnalysis* Chichester: Wiley.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. (2010). *Metaanalysis: Computing treatment effects and effect sizes*. Chichester: Wiley.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. (2011). *Comprehensive MetaAnalysis Version 2,* Biostat. Găsit la www.MetaAnalysis.com.

Clark-Carter, D. (2010), *Quantitative psychological research: the complete student's companion* (ediţia a III-a). New York: Psychology Press.

Campbell Collaboration (2012). *What is systematic review?* Disponibil la adresa http://www.campbellcollaboration.org/

Cochrane Collaboration (2011). *RevMan*. Găsit la adresa www.cc-ims.net/revman.

Cohn, L.D., Becker, L.A. (2003). How metaanalysis increases statistical power. *Psychological Methods*, 8, pp. 243-253.

Cooper, H. (2004). *Research synthesis and metaanalysis. A step by step approach* (ediţia a IV-a). Thousand Oaks: Sage.

Costa, P.T., McRae, R.R. (1992). *Revised NEO personality inventory and NEO-five-factor inventory: Professional manual*. Odessa FL, USA: Psychological Assessment Resources.

David, D., Szentagotai, A., Lupu, V., Cosman, D. (2008). Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: A randomized clinical trial, post-treatment outcomes, and six-month follow-up. *Journal of Clinical Psychology*, 64, pp. 728-746.

DeCoster, J. (2004). *Metaanalysis notes*. Găsit la adresa http://www.stat-help.com/notes.html.

DeCoster, J. (2011). *Microsoft Excel Spreadsheets*. Disponibil la adresa http://www.stat-help.com/spreadsheets.html.

Duval, S., Tweedie, R. (2000). A nonparametric „trim and fill" method of accounting for publication bias in metaanalysis. *Journal of the American Statistical Association*, 95, pp. 89-99.

Egger, M., Davey Smith, G., Schneider, M; Minder, C. (1997). Bias in metaanalysis detected by a simple, graphical test. *British Medical Journal*, 315, pp. 629-634.

Field, A. (2001). Metaanalysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, pp. 161-180.

Field, A. (2003). Can metaanalysis be trusted? *The Psychologist*, 16, pp. 642-645.

Field, A. (2005). Metaanalysis. În J. Miles şi P. Gilbert (eds.), *A Handbook of Research Methods in Clinical & Health Psychology* (pp. 295-308). Oxford: Oxford University Press.

Green, S.G., McDonald, S. (2005). The Cochrane Collaboration: More than systematic reviews? *Internal Medicine Journal*, 35, pp. 4-5.

Hedges, L.V., Olkin, I. (1985). *Statistical methods for metaanalysis*. Orlando: Academic Press.

Higgins J.P.T., Green S (eds.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Disponibil la:

www.cochrane-handbook.org.

Higgins, J.P.T., Thompson, S.G.. Deeks, J.J., Altman, D.G. (2003). Measuring inconsistency in metaanalyses. *British Medical Journal*, 327, pp. 557-560.

Hunter, J.E., Schmidt, F.L. (2004). *Methods of MetaAnalysis: Correcting Error and Bias in Research Findings* (ediţia a II-a). Thousand Oaks: Sage.

Linden, W., Phillips, M.J., Leclerc, J. (2007). Psychological treatment of cardiac patients: A metaanalysis. *European Heart Journal*, 28, pp. 2972-2984.

Lipsey, M.W., Wilson, D.B. (2001). *Practical Metaanalysis*. Thousand Oaks: Sage.

Lyons, L.C., Morris, W.A. (2012). *The MetaAnalysis Calculator*. Disponibil la adresa http://www.lyonsmorris.com/ma1/index.cfm

Maricuţoiu, L.P. (2008). Introducere în metaanaliză. Metaanaliza corelaţiilor. *Psihologia Resurselor Umane*, 6, pp. 112-127.

Maricuţoiu, L.P., Sava, F.A. (2007). *PowerStaTim 1.0 – Analiza puterii statistice şi a mărimii efectului*. Disponibil la: www.psihologietm.ro

McDaniel, M.A. (2005). Big-brained people are smarter: A metaanalysis of the relationship between in vivo brain volume and intelligence. *Intelligence*, 33, pp. 337-346.

Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P., Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16, pp. 62-73.

Moher, D., Jadad, A.R., Tugwell, P. (1996). Assessing the quality of randomized controlled trials: Current issues and future directions. *International Journal of Technology Assessment in Health Care*, 12, pp. 195-208

Morris, S.B., DeShon, R.P. (2002). Combining effect size estimates in metaanalysis with repeated measures and independent-groups designs. *Psychological Methods*, 7, pp. 105-125.

Mulrow, C.D. (1994). Systematic reviews: Rationale for systematic reviews. *British Medical Journal*, 309, pp. 597-599.

Olejnik, S., Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8, pp. 434-447.

Radu, I., Miclea, M., Albu, M., Moldovan, O., Nemeş, S., Szamoskozy, S. (1993). *Metodologie psihologica si analiza datelor*. Bucureşti : Sincron.

Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. *Journal of Clinical Psychology*, 63, pp. 611-631. US Department of Education, Office of Planning, Evaluation, and Policy Development (2010). *Evaluation of evidence-based practices in online learning: A metaanalysis and review of online learning studies*. Washington DC, disponibil la adresa www.ed.gov/about/offices/list/opepd/ppss/reports.html

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., Tugwell, P. (2008). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in*

metaanalyses. disponibil la adresa: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm

Wikipedia (2012). *List of academic databases and search engines*. disponibil la adresa: http://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines

1. Dintre sintezele de tip narativ, cel mai întâlnit tip este sinteza teoretică (Cooper, 2004). Aceasta are rolul de a evalua comparativ teoriile utilizate pentru a vedea care dintre ele explică mai bine anumite fenomene, punând accent pe descrierea unor experimente critice pentru a diferenţia între teorii, pe dovezile empirice care contrazic sau sprijină anumite modele teoretice şi pe anumite reformulări/modificări conceptuale necesare pentru a nu respinge o anumită teorie (revizuiri ale teoriilor).

2. Conform Campbell Collaboration (2012), o sinteză a literaturii de specialitate poate fi denumită evaluare sistematică dacă îndeplineşte următoarele criterii: (a) se bazează pe criterii clare de includere/excludere a studiilor; (b) specifică explicit strategia utilizată pentru căutarea studiilor; (c) foloseşte un sistem de codare a variabilelor de interes din studiile selectate în baza unui manual de codare; (d) se finalizează printr-o metaanaliză (dacă este posibil).

3. Pentru o listă extinsă a bazelor de date internaţionale existente, recomand consultarea Wikipedia (2012).

4. Maricuţoiu (2008) introduce pentru publicul de la noi această formă de metaanaliză, discutând cazul particular al metaanalizei corelaţiilor.

5. Printre cei care au colaborat la realizarea acestui program se regăsesc: David Wilson, Betsy Becker, Julian Higgins, Will Shadish, Hannah Rothstein, Michael Borenstein, Larry Hedges, Harris Cooper şi Mark Lipsey.

6. O excepţie în care putem lua în analiză mai mulţi indicatori output proveniţi din acelaşi studiu este cazul în care suntem interesaţi să urmărim în cadrul metaanalizei dacă instrumentele de lucru constituie o variabilă moderatoare (de exemplu, dacă efectul terapiei asupra simptomatologiei depresive este moderat de tipul de instrument utilizat – Beck sau Hamilton). Pentru a răspunde punctual la acest aspect, vom lucra nu cu un eşantion de studii, ci cu un eşantion de mărimi ale efectelor, fapt ce ne va permite să testăm dacă există diferenţe semnificative de mărime a efectului între studiile care au utilizat proba Hamilton, respectiv proba Beck. O altă excepţie este cazul în care urmărim efecte pe paliere diferite. De exemplu, într-un studiu care vizează efectul jocurilor video cu conţinut violent asupra agresivităţii la copii pot fi analizate separat efectele produse la nivelul cogniţiilor agresive şi cele care vizează comportamentul agresiv.

7. Field (2001), în baza unei simulări de tip Monte Carlo, concluzionează că în cazul metaanalizei efectelor random aplicate în cazul corelaţiilor, atât pentru metoda descrisă (Borenstein *et al.*, 2009, bazată pe Hedges şi Olkin, 1985), cât şi pentru metoda propusă de Hunter şi Schmidt (2004), apare un risc crescut de eroare de tip I dacă sunt utilizate cel mult 15 studii.

8. Studiile metaanalitice, comparativ cu cele empirice (primare), au o putere statistică mult mai mare (îndeosebi metaanaliza efectelor fixe), crescând şansele de a găsi efecte semnificative statistic. Sugestiv ar fi exemplul amintit de Cohn şi Becker (2003), al unei metaanalize ce vizează impactul unui tratament medicamentos aplicat pacienţilor care au suferit un infarct miocardic. Deşi în 27 dintre cele 33 de studii incluse în metaanaliză au fost observate efecte nesemnificative statistic, efectul mediu rezultat din metaanaliză a fost unul semnificativ statistic, scăzând cu 20% riscul mortalităţii în rândul pacienţilor care au beneficiat de tratamentul respectiv.

9. După unii autori, precum Field (2003), a cunoaşte aceşti factori moderatori este mai important decât rezultatele generale (media ponderată a mărimii efectului) în sine. În această direcţie, Field (2003) apelează la o parabolă elocventă, aceea a realizării unei metaanalize pe baza a trei studii, efectuate pe un număr egal de participanţi, în care s-au obţinut următoarele efecte ale unui medicament miraculos care să-i ajute pe studenţi să înţeleagă cunoştinţele de statistică: d = 0,45 (în studiul din SUA), d = 0,00 (în studiul din Belgia) şi d = –0,45 (în studiul din Marea Britanie). Cititorii neutri ar putea concluziona că medicamentul este unul ineficient, deoarece media celor trei studii este 0. Însă această concluzie ar fi valabilă doar pentru belgieni, deoarece în SUA medicamentul a funcţionat, iar în Marea Britanie a avut efecte adverse, contrar celor aşteptate. De aceea, mai important decât media în sine este să vedem care e nivelul de omogenitate/eterogenitate al efectului de-a lungul studiilor şi să găsim care sunt acele variabile moderatoare responsabile pentru variaţiile observate în mărimea efectului.

Capitolul 8

Raportarea rezultatelor în articole de specialitate

Calitatea unei cercetări poate fi subminată sau întărită de capacitatea autorilor de a aşterne în scris rezultatele cercetărilor lor. Forma consacrată de comunicare a acestor rezultate este cea a articolelor de specialitate, supuse procesului de evaluare colegială (*peer-review papers*). Dacă articolul este coerent şi bine structurat, le va fi mai uşor atât evaluatorilor, cât şi publicului cititor să înţeleagă aspectele importante ale studiului, crescând şansele publicării, respectiv şansele ca articolul să aibă un impact mai ridicat printre specialişti. Mai mult, capacitatea autorilor de a scrie oferă adesea indicii cu privire la maturitatea lor, cazurile de articole scrise foarte clar fiind mult mai frecvente printre cercetătorii cu experienţă decât printre cercetătorii tineri.

Ţinând cont de aceste considerente, în acest capitol vom prezenta o serie de recomandări privind structurarea unui articol de specialitate, dar şi privind procesul editorial asociat acestui demers. La fel ca în cazul celorlalte capitole, accentul va fi pus de aspectele aplicative, pragmatice, pentru a-i ajuta pe cei interesaţi să valorizeze la maximum potenţialul cercetărilor realizate sau în curs de realizare. Ne vom referi îndeosebi la articolele empirice (cele care presupun colectarea şi analiza unor date de către cercetători), dar vom face unele referiri şi la alte categorii de articole cum sunt cele din zona studiilor de sinteză: studiile metaanalitice şi

sintezele narative (teoretice). O altă precizare necesară este aceea că ne vom focaliza asupra structurării articolelor adresate unui public-ţintă cunoscător (reviste cu nivel ridicat de specializare) şi mai puţin asupra publicării unor articole în reviste de popularizare[1].

8.1. Anatomia unui articol empiric

Acest tip de studiu este, de departe, cel mai răspândit tip de articol ştiinţific, deoarece sub această formă sunt cuprinse toate studiile primare, bazate pe colectarea şi analiza datelor.

Datorită răspândirii acestui tip de studiu, structura articolelor empirice pare familiară atât specialiştilor, cât şi cititorilor aflaţi în perioada de formare profesională (studenţi, masteranzi), ea putând fi înţeleasă, într-o formă simplă, ca o structură în patru paşi: introducere („Care este problema investigată?"), metodologia cercetării („Cum a fost investigată problema?"), rezultate („Ce rezultate s-au găsit?") şi discuţii („Ce implicaţii/relevanţă au aceste rezultate?") (Hall, 2008)[2].

Totuşi, într-o formă extinsă am putea vorbi de opt sau chiar nouă secţiuni ale unui articol empiric în forma sa publicată (Cooper, 2011): titlu, nota autorului, rezumat (*abstract*), introducere, metodologia cercetării, rezultate, discuţii, bibliografie şi, eventual, anexe (materiale suplimentare disponibile online). Asupra fiecăruia dintre aceste aspecte ne vom opri în continuare, făcând o serie de recomandări specifice.

8.1.1. Titlul articolului

Alegerea titlului pentru articolele din domeniul psihologiei ar trebui să ţină seama atât de recomandările din domeniu (de exemplu, APA recomandă ca titlul să nu depăşească 12 cuvinte[3], fiind preferate formulările concise; APA, 2010), cât şi de principalele trei calităţi pe care ar trebui să le aibă un titlu. Aceste calităţi sunt: a fi informativ (să nu inducă în eroare), a fi atractiv (să trezească interesul cititorilor) şi a include cuvinte-cheie (să crească şansele de identificare ulterioară a articolului de către cei care vor căuta articole pe tema respectivă în baze de date internaţionale (Moss, 2008).

Din perspectiva formei titlului, în domeniul psihologiei sunt utilizate frecvent două tipuri de titluri: nominale şi compuse (Budgell, 2009). În categoria titlurilor nominale se regăsesc acele titluri formate dintr-o afirmaţie care conţine ideea principală a articolului. Titlurile compuse sunt formate din două părţi. Prima parte poate fi o întrebare, poate fi un artificiu retoric sau o descriere succintă a aspectelor-cheie din studiu, în timp ce în partea a doua sunt oferite precizări suplimentare cu privire la tipul studiului, la alte caracteristici metodologice esenţiale ale studiului este.

Caseta 8.1. *Exemple de titluri*

Exemple de titluri nominale[4]

1. Efectul terapiei familial-sistemice asupra adolescenţilor cu depresie majoră
2. Relaţia dintre angajamentul organizaţional şi fluctuaţia de personal a cadrelor didactice
3. Inteligenţa generală ca predictor al performanţei extra-rol la angajaţii din industria uşoară
4. Impactul grădiniţei cu program prelungit asupra achiziţiilor educaţionale şi a stării de bine a preşcolarilor

Exemple de titluri compuse

Descriere succintă + detaliu metodologic

1. Tipuri de personalitate în baza modelului Big Five: o analiză de cluster pe populaţia din România.
2. Efectul terapiei raţional-emotive în reducerea cogniţiilor iraţionale la delincvenţi: un studiu clinic controlat
3. Relaţia dintre personalitate şi absenteismul de la locul de muncă: o metaanaliză.

Artificiu retoric + descriere succintă (+ detaliu metodologic)

4. Noi credem în moderaţie: personalitate şi abilităţi de muncă în echipă[5]
5. Partea întunecată a predării: o metaanaliză a impactului comportamentelor didactogene ale profesorului asupra elevilor[6]

Întrebare + detaliu metodologic (alte aspecte)

6. Sunt eficiente programele de training pentru părinţii copiilor cu ADHD? O metaanaliză
7. Tutoriate online *vs* tutoriate faţă în faţă: care sunt mai eficiente? Un studiu longitudinal

După cum se poate constata din caseta 8.1, titlurile nominale sunt cele mai succinte şi mai apropiate de esenţa problemei investigate, dacă ar fi să ne raportăm la modelul PI(C)O de formulare a problemei de cercetat. Pe de altă parte, asemenea formulări tind să atragă mai puţin atenţia (cel puţin comparativ cu formulările retorice) şi să fie mai puţin utile pentru cercetătorii interesaţi să distingă încă din titlu despre ce fel de studiu este vorba (de exemplu, o metaanaliză, un studiu randomizat, un studiu longitudinal etc.).

8.1.2. Nota autorului

Scopul acestei mini-secţiuni (*author note*) este de a prezenta autorii articolului şi eventualele circumstanţe speciale care ar trebui cunoscute de către publicul cititor. Standardele APA (2010) recomandă standardizarea acestui proces, prin cuprinderea următoarelor informaţii: (a) numele autorului (autorilor) şi afilierea instituţională şi departamentală a acestora la momentul derulării studiului, respectiv în prezent (numai pentru autorii care şi-au modificat între timp afilierea instituţională); (b) mulţumiri aduse (pentru sprijinul financiar primit în derularea studiului – numărul *grantului*, instituţia finanţatoare etc. – respectiv pentru contribuţia adusă la reuşita studiului de către persoane care nu sunt incluse în lista de autori); (c) alte precizări explicite necesare în cazuri speciale (precizarea situaţiilor în care există un conflict de interese; menţionarea eventualelor studii publicate anterior care se bazează pe acelaşi eşantion sau aceeaşi bază de date; specificarea faptului că ordinea autorilor nu reflectă contribuţia lor la articol, toţi având contribuţii egale etc.); (d) date ale persoanei de contact.

Caseta 8.2. *Exemplu de notă a autorului*

Adaptare după Sava, Yates, Lupu, Szentagotai şi David (2009)

Florin A. Sava, Departamentul de Psihologie Clinică şi Psihoterapie, Universitatea „Babeş-Bolyai” din Cluj-Napoca; Brian T. Yates, Departamentul de Psihologie, American University, Washington DC; Viorel Lupu, Departamentul de Fiziologie, UMF „Iuliu Haţeganu” din Cluj-Napoca; Aurora Szentagotai, Departamentul de Psihologie, Universitatea „Babeş-Bolyai” din Cluj-Napoca; Daniel David, Departamentul de Psihologie Clinică şi Psihoterapie, Universitatea „Babeş-Bolyai” din Cluj-Napoca.
Florin A. Sava este afiliat în prezent la Departamentul de Psihologie, Universitatea de Vest din Timişoara.

Mulţumiri: această cercetare a fost susţinută de Institutul Albert Ellis, de CNCS şi de Centrul Român de Psihoterapie Cognitiv-Comportamentală, prin granturile 113, 33374, respectiv 9).

Aceasta publicaţie se bazează pe aceeaşi bază de date utilizată în articolul menţionat mai jos. Dacă în articolul respectiv s-a urmărit exclusiv compararea eficienţei celor trei intervenţii în combaterea depresiei majore, acest articol abordează comparativ raporturile cost-eficienţă şi cost-utilitate pentru cele trei intervenţii.
David, D., Szentagotai, A., Lupu, V., & Cosman, D. (2008). „Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: A randomized clinical trial, posttreatment outcomes, and six-month follow-up", *Journal of Clinical Psychology*, 64, 728-746.

Date de contact: corespondenţa pentru acest articol ar trebui trimisă către Florin A. Sava, Departamentul de Psihologie, Universitatea de Vest din Timişoara, Bd. Vasile Pârvan nr. 4, 300223, Timişoara, România, e-mail: afsava@socio.uvt.ro. Pentru autorii care şi-au schimbat afilierea din momentul realizării studiului până în momentul publicării acestuia, cutuma este să se treacă ca afiliere instituţia de care autorul aparţine în momentul realizării studiului.

8.1.3. Rezumatul articolului şi cuvintele-cheie

Modul de redactare a rezumatului unui articol este dependent, într-o anumită măsură, de recomandările revistei în care se intenţionează a fi publicat articolul, fiind întâlnite formate mai scurte (până la 150 de cuvinte) sau mai lungi (până la 200-250 de cuvinte). Primele sunt scrise, de regulă, sub forma unui singur paragraf (rezumate narative), celelalte sub formă structurată în câteva mici secţiuni (rezumate structurate) (Budgell, 2009). În caseta 8.3 este prezentat câte un exemplu pentru fiecare dintre cele două tipuri de rezumate, precum şi un exemplu de rezumat pentru cazul articolelor cu studii multiple. Indiferent de formatul de redactare, formatul JARS[7] recomandă includerea în rezumat doar a datelor de tip factual (fără aprecieri sau judecăţi de valoare), care să facă referire la: (a) problema investigată (*background* şi/sau obiectivul studiului); (b) detalii despre participanţi; (c) detalii-cheie despre metodologia studiului; (d) rezultate; (e) concluzii/implicaţii/aplicaţii.

Desigur, acest gen de raportare este funcţional îndeosebi atunci când articolul respectiv raportează un singur studiu. Pentru articolele care conţin mai multe studii, nivelul de abstractizare creşte.

Caseta 8.3. *Exemple de rezumate*

Adaptare după Sava, Maricuţoiu, Rusu, Macsinga şi Vîrgă (2011) – 136 de cuvinte

Relaţia dintre credinţele iraţionale şi evaluările explicite, respectiv implicite ale stimei de sine a fost investigată în două studii corelaţionale independente (N1 = 117; N2 = 102), efectuate pe studenţi de la nivelul ciclului de licenţă. În ambele studii cogniţiile iraţionale şi stima de sine explicită au fost evaluate prin GABS-SV, respectiv scala Rosenberg, în timp ce stima de sine implicită a fost evaluată în primul studiu prin testul asocierilor implicite (IAT), iar în al doilea studiu prin proba iniţialelor numelui. Două rezultate robuste (replicate) indică existenţa unei corelaţii negative între stima de sine explicită şi autoblamare, o credinţă iraţională specifică, precum şi absenţa unei corelaţii între nivelul stimei de sine implicite şi oricare dintre credinţele iraţionale. Aceste rezultate sugerează ideea că disputarea cogniţiilor iraţionale în cadrul intervenţiilor psihoterapeutice afectează doar cogniţiile explicite, cele implicite rămânând nealterate.

Adaptare după Sava, Yates, Lupu, Szentagotai şi David (2009) – 198 de cuvinte

Obiectiv. Studiul compară raportul cost-eficienţă în cazul utilizării terapiei cognitiv-comportamentale (TCC), respectiv al fluoxetinei (Prozac) pentru tratamentul depresiei majore într-un studiu clinic controlat.
Participanţi şi design. 110 adolescenţi români nepsihotici, aflaţi la primul episod diagnosticat de depresie majoră, au fost distribuiţi randomizat în două grupe: TCC (N = 56) şi fluoxetină (N = 54). Ambele intervenţii s-au derulat săptămânal, timp de 14 săptămâni.
Variabilele dependente principale. Nivelul depresiei a fost evaluat în trei momente diferite (pretest, posttest şi *follow-up* la şase luni distanţă) prin scala de depresie Beck.
Rezultate. Ambele forme de intervenţie reduc semnificativ statistic nivelul depresiei, raportându-ne la nivelul existent la pretest. Nu există diferenţe semnificative între TCC şi fluoxetină cu privire la eficienţa lor în combaterea depresiei majore (d Cohen = 0,07 [–0,13; 0,22] la nivel de posttest, respectiv 0,02 [–0,12; 0,15] la nivel de *follow-up*). În schimb, TCC a obţinut un raport cost-eficienţă mai bun decât fluoxetina, datorită costurilor mai scăzute: mediana 26,44 USD/zi fără depresie pentru TCC, respectiv 34,93 USD/zi fără depresie pentru fluoxetină.

Implicaţii. Rezultatul raportului cost-eficienţă este dependent de contextul economic, costul serviciilor de psihoterapie fiind mai scăzut decât costul farmacoterapiei în ţări cu economii emergente[8].

Din perspectiva utilităţii unui rezumat, acesta îndeplineşte două funcţii majore: (a) să prezinte clar, într-o manieră concisă, ideea şi rezultatele unui studiu, fiind primul şi, de multe ori, singurul paragraf citit dintr-o cercetare (aspect ilustrat în caseta 8.3); (b) să crească şansele ca articolul să fie citat, prin facilitarea regăsirii acestuia în baze de date internaţionale.

Cu privire la cea de-a doua perspectivă, voi prezenta două recomandări utile:

- sunt preferabile formulările cu un nivel redus de abstractizare; de exemplu, în loc să se spună că este vorba de 110 participanţi, este preferabil să spunem că este vorba de 110 adolescenţi din România; acest lucru va ajuta la identificarea articolului în baze de date internaţionale (de exemplu, cineva interesat să identifice eficienţa diferitelor forme de terapie pentru adolescenţi va introduce, foarte probabil, între cuvintele-cheie în limba engleză termeni precum *teenagers* sau *adolescents*; dacă în rezumat vom trece cuvântul participanţi şi dacă titlul nu conţine referiri exacte la populaţia-ţintă, articolul nostru nu va fi selectat pe baza celor două cuvinte specificate; reversul nu este însă valabil deoarece dacă cineva este interesat de nivelul eficienţei terapiei la nivel global, e puţin probabil să introducă termeni generici, precum *participants*, preferând să se focalizeze pe alte cuvinte-cheie; aceeaşi recomandare de a utiliza termeni categoriali cât mai concreţi este valabilă şi pentru alte aspecte, precum tipul de studiu sau instrumentele utilizate; de exemplu, ar fi dezirabil să se facă referire la tipul de studiu, astfel încît în locul unor

exprimări de genul *in this study* ar fi mai bine să fie precizat exact tipul studiului (*in this correlational study*; *in this randomized trial*) pentru a uşura munca de selectare a studiilor relevante atunci când alţi autori sunt interesaţi de o anumită tipologie de studii;

- dacă rezumatele sunt generoase ca număr de cuvinte permise, este recomandabil să fie prezentate explicit rezultatele obţinute astfel încât să poată fi valorificate rezultatele articolului în metaanalize (vezi al doilea exemplu din caseta 8.3); acest lucru este util deoarece multe baze de date indexează doar rezumatul articolelor, dar chiar şi pentru articolele indexate integral există multe situaţii în care accesul la acestea se face doar în regim contra cost.

Desigur, formatele pentru rezumat prezentate anterior se referă la cazul studiilor empirice. Pentru alte tipuri de studii, cum sunt evaluările sistematice, este preferabil un model structurat cu şase puncte (Budgell, 2009) – obiectiv, sursa datelor, selecţia studiilor, metode de extragere a datelor, rezultate sintetice şi concluzii –, precum cel prezentat în caseta 8.4. Detalii despre alte tipuri de studii (metodologice, teoretice, studii de caz etc.) se găsesc în manualul oficial al standardelor de publicare APA (2010).

Caseta 8.4. *Exemplu de rezumat pentru o metaanaliză*

Adaptare după Szentagotai şi David (2010) – 214 cuvinte

Obiectiv. Scopul studiului este de a realiza o metaanaliză cu referire la efectul terapiei cognitiv-comportamentale (TCC) ca tratament adjuvant la medicaţie în cazul pacienţilor diagnosticaţi cu depresie bipolară.
Sursa datelor. Au fost incluse articolele publicate în perioada 1980-2008, selectate din baza de date Medline. Cuvintele-cheie utilizate au fost: cognitiv şi tulburare bipolară, terapie cognitivă şi tulburare bipolară, terapie cognitiv-comportamentală şi tulburare bipolară.

Selecţia studiilor. Criteriile de includere se referă la: (a) studii clinice controlate în care una dintre intervenţii este TCC; (b) existenţa unui grup de control care primeşte doar medicaţie; (c) existenţa datelor suficiente pentru calculul mărimii efectului.
Metode de extragere. Indicatorul d Cohen a fost utilizat pentru calcularea mărimii efectelor, acestea fiind reunite apelând la modelul metaanalizei efectelor aleatorii.
Rezultate. A fost obţinut un efect scăzut spre moderat al impactului adjuvant al TCC la nivel de posttest ($d = 0,42$, $p < 0,05$), respectiv de *follow-up* ($d = 0,27$, $p < 0,05$). Efecte similare au fost obţinute cu privire şi la calitatea vieţii ($d = 0,36$, $p < 0,05$) şi la complianţa la tratament ($d = 0,53$, $p < 0,05$).
Concluzii. TCC are o eficienţă moderată ca tratament adjuvant pentru farmacoterapie în cazul tulburărilor bipolare, însă sunt necesare strategii noi de lucru în TCC care să intensifice efectul benefic al psihoterapiei.

În privinţa cuvintelor-cheie ce însoţesc un rezumat, strategia de lucru este uşor diferită, fiind preferabili termenii-cheie cu frecvenţa cea mai ridicată, ceea ce implică adesea un nivel de generalizare mai ridicat. Astfel, foarte utilă ar fi selectarea unor cuvinte cu frecvenţă de apariţie ridicată, cum ar fi termenii-cheie incluşi în tezaurul de cuvinte al unor baze de date internaţionale. De exemplu, dacă cineva realizează un studiu privind relaţia profesor-elev, ar putea utiliza termenul-cheie *teacher-student interaction*, aceasta fiind denumirea întâlnită în tezaurul de termeni din baza de date PsycINFO (domeniul psihologiei) sau *teacher-student relationship*, dacă se are în vedere tezaurul de termeni din baza de date ERIC (domeniul ştiinţelor educaţiei). De asemenea, dacă se are în vedere un studiu care să identifice trăsături de personalitate precum conştiinciozitatea sau stabilitatea emoţională, un termen potrivit ar fi *personality* sau *five-factor personality model* – categorii superioare ca nivel de abstractizare.

De asemenea, o altă recomandare ar fi ca între termenii-cheie ai unui articol să se regăsească termeni relevanţi pentru a descrie obiectivul studiului. De exemplu, pentru studiul efectuat de Szentagotai şi David

(2010) am putea include următoarele cuvinte-cheie în limba engleză: *metaanalysis*, *cognitive-behavioral therapy*, *pharmacotherapy*, *bipolar disorder*, *incremental effect*, în timp ce în studiul realizat de Sava *et al.* (2009) am putea include, pornind de la obiectivul formulat, termeni în limba engleză precum: *cost-effectiveness*, *cognitive-behavioral therapy*, *pharmacotherapy*, *depression*, *adolescents*, *randomized trial*.

8.1.4. Introducerea (partea introductivă dintr-un studiu empiric)

Cei mai mulţi specialişti formaţi sau în curs de formare (studenţi, masteranzi, doctoranzi) nu au abilitatea de a scrie bine partea introductivă a unui articol. La acest fapt contribuie: (a) absenţa precizării explicite a unei structuri în redactarea introducerii spre deosebire de o altă secţiune a studiului – metodologia cercetării (unde subcomponentele sunt prezentate explicit prin termeni precum *participanţi*, *instrumente*, *procedură*); (b) interferenţa cu modele de redactare a introducerii specifice unor texte mai lungi (lucrări de disertaţie, teze de doctorat).

Pe ansamblu, introducerea unui articol empiric are două mari părţi: motivarea studiului (*study rationale*) şi scopul studiului (*study purpose*) (Girden, 2001). Din perspectiva unui proces de rezolvare a problemelor, în cadrul motivării studiului este prezentată problema, încercările anterioare de rezolvare, necunoscutele existente şi impactul acestora asupra stării de fapt, în timp ce în partea destinată scopului studiului este prezentată soluţia avută în vedere şi este oferită o justificare pentru alegerea acesteia. Pentru a concretiza acest demers explicativ, vom apela la exemplul unui studiu ce vizează evaluarea implicită a personalităţii

(Sava *et al.*, 2012); vom ilustra astfel toate componentele aferente unui studiu empiric, în general.

Tabelul 8.1. *Descrierea componentelor tipice ale unei secţiuni introductive, rolul lor şi eventuale recomandări. Ilustrarea cazului evaluării implicite a personalităţii*

Subsecţiune	Rol şi recomandări	Exemplificare (formă simplificată prin eliminarea celor mai multe referinţe bibliografice şi a unor porţiuni de text)
Motivarea studiului (argumentarea necesităţii studiului) – De ce ar fi nevoie de acest studiu?		
1. Introducerea problemei *Ce investigăm?*	Are rolul de a capta atenţia asupra problemei ce urmează a fi investigată, subliniind importanţa temei din punct de vedere teoretic sau practic. *Recomandări*: utilizaţi statistici sau afirmaţii de impact, care să atragă atenţia; utilizaţi timpul prezent. „Între 1990 şi 2010, consumul de droguri a crescut…”	Evaluarea personalităţii se bazează aproape în exclusivitate pe măsuri explicite, autoraportate, de tipul inventarelor de personalitate. Acestea prezintă însă două limite importante: oferirea de răspunsuri dezirabile social (Furnham, 1997)[9] , respectiv existenţa limitelor introspective referitoare la o cunoaştere corectă de sine (Paulhus, Reid, 1991). Cercetările din domeniul cogniţiei sociale din ultimul deceniu susţin faptul că metodele implicite de evaluare pot constitui alternative viabile pentru măsurarea unor constructe psihologice, fără a face apel la autoraportare (Gawronski, Payne, 2010). Până în prezent, cele mai multe studii de acest gen au fost efectuate pentru a măsura atitudinile (Banaji, 1995), metodele implicite fiind mai puţin utilizate pentru evaluarea personalităţii.
2. Prezentarea relevanţei problemei	Se justifică implicaţiile teoretice şi/sau practice ale problemei.	Aşa cum utilizarea măsurilor implicite pentru evaluarea atitudinilor a condus la dezvoltări teoretice importante (vezi modelul dual al lui Gawronski şi

De ce este important să investigăm acest lucru?		Bodenhausen, 2006), este de aşteptat să fie dezvoltate noi modele teoretice în urma utilizării metodelor implicite de investigare a personalităţii. De asemenea, dintr-o perspectivă pragmatică, dovezile privind rezistenţa metodelor implicite la încercarea de a oferi răspunsuri dezirabile social sunt de interes pentru practicienii din domeniul selecţiei de personal. Astfel, studiul metodelor implicite de investigare a personalităţii se poate dovedi util, atât teoretic, cât şi aplicativ.
3. Studii anterioare relevante *Ce s-a făcut până acum cu privire la problema investigată?*	Rolul acestora este de a prezenta ce eforturi au fost făcute până în prezent cu referire la tematica studiată. *Recomandări*: accent direct pe studii relevante pentru tematica abordată. Este nerecomandată o abordare secvenţială de genul: „Ce sunt metodele implicite?”, „Ce este personalitatea?” şi abia apoi aspectul relevant – relaţia dintre metodele implicite şi personalitate. Este recomandabil să se distingă la nivelul verbelor între rezultatele („autorii au demonstrat/au identificat/au arătat că” etc.) şi presupoziţiile anterioare („autorii au afirmat că/sunt de părere/consideră că” etc.).	O serie de studii care au folosit testul asocierilor implicite (IAT) ca metodă implicită de evaluare a personalităţii au arătat că tehnica IAT posedă o consistenţă internă ridicată (Gawronski, 2009) şi un nivel acceptabil al validităţii relative la criteriu (Perugini şi Richetin, 2010). Tehnica a fost utilizată pentru a surprinde atât cele cinci trăsături de personalitate ale modelului Big Five (Smuckle *et al.*, 2008), cât şi alte trăsături particulare, precum anxietatea sau nivelul de agresivitate. Au existat, de asemenea, încercări de a evalua personalitatea prin alte metode implicite, precum Procedura Asocierii Implicite (IAP) (Schnabel *et al.*, 2006), sarcina afectivă extrinsecă Simon (EAST) (Teige *et al.*, 2004) sau sarcina de asociere Go/No Go (GNAT) (Boldero *et al.*, 2007).
4. Limitele studiilor anterioare[10]	Rolul acestui calup de afirmaţii este de a pregăti terenul pentru	Rezultatele obţinute în urma utilizării acestor tehnici sunt mixte. Astfel, în

Care sunt neajunsurile studiilor anterioare referitor la problema investigată?	soluţia propusă prin sublinierea neajunsurilor încercărilor anterioare. *Recomandări*: focalizaţi-vă pe lipsa de informaţii („nu se cunoaşte rolul…”; „nu s-au centrat pe…”); pe surprinderea unor rezultate contradictorii („nu este clar rolul..”) sau asupra limitelor acestora relevante pentru studiul realizat („studiile anterioare fiind demersuri corelaţionale nu pot stabili o relaţie de tip cauză-efect”; „nu a fost luat în considerare efectul moderator al…” etc.)	cazul tehnicilor IAP, EAST şi GNAT s-a observat că acestea au proprietăţi psihometrice inadecvate pentru a putea fi utilizate ca instrumente de evaluare într-un domeniu sensibil, cum este cel al personalităţii. Principalele neajunsuri ale acestor tehnici constau în niveluri scăzute ale consistenţei interne (variind între 0,30 şi 0,50), precum şi în niveluri ridicate de varianţă comună, ceea ce conduce la corelaţii nejustificat de mari între două trăsături de personalitate măsurate cu aceeaşi probă. În schimb, rezultate ceva mai bune au fost obţinute cu ajutorul tehnicii IAT, evidenţiindu-se niveluri ridicate ale consistenţei interne (valoarea mediană fiind 0,76). Totuşi, inclusiv în cazul IAT au fost observate valori scăzute ale coeficientului de stabilitate test-retest, ceea ce ridică semne de întrebare cu referire la efectul măsurat, trăsăturile de personalitate fiind considerate constructe stabile în timp, fapt ce ar fi trebuit să se reflecte prin valori ridicate ale corelaţiei dintre test şi retest.
Scopul studiului		
5. Precizarea obiectivului *Ce se urmăreşte în acest studiu şi ce aduce el în plus?*	Rolul acestei secţiuni este de a introduce (prezenta) soluţia alternativă/adică ceea ce se urmăreşte prin studiul de faţă, demers însoţit de o justificare a alegerii făcute. *Recomandări*: ca formulare se revine la timpul prezent şi se oferă o motivare pentru obiectivul ales (de ce a fost ales) şi pentru utilitatea acestuia	În lucrarea de faţă vom urmări în ce măsură poate fi evaluată la nivel implicit personalitatea prin apelul la o altă tehnică din care este derivată, care nu a fost utilizată până în prezent în scopul evaluării personalităţii, denumită atribuire eronată a conţinutului semantic (SMP – *semantic misattribution procedure*). Tehnica este derivată din atribuirea eronată a afectului (AMP), dezvoltată de Payne

	(prin ce anume contribuie la progresul cunoaşterii)[11].	*et al.* (2005). Am ales această tehnică deoarece studiile anterioare referitoare la atitudinile implicite au arătat că posedă proprietăţi psihometrice foarte bune, fiind, alături de IAT, una dintre cele mai robuste tehnici implicite (Gawronski, 2009). De asemenea, s-a demonstrat că tehnica din care este derivată este rezistentă la încercări de a oferi răspunsuri dezirabile… etc. (se adaugă un paragraf cu descrierea tehnicii şi o justificare a mecanismului psihologic prin care tehnica ar putea funcţiona pentru a evalua personalitatea). De aceea, demonstrarea posibilităţii de a surprinde personalitatea printr-o metodă implicită ar constitui un progres în evaluarea personalităţii, mai ales în contexte în care indivizii sunt predispuşi să ofere răspunsuri dezirabile social.
6. Specificarea ipotezelor şi operaţionalizare *Ce strategie s-a utilizat pentru atingerea obiectivului propus?*	Această parte încheie secţiunea introductivă. Sunt prezentate aici ipotezele, inclusiv o explicitare a demersului teoretic (argumentativ) care a condus la aceste ipoteze, precum şi scurte referiri la designul cercetării şi la variabilele incluse pentru a răspunde la ipotezele respective. O situaţie aparte o reprezintă articolele cu studii multiple, caz în care în această secvenţă se explicitează mai degrabă rolul fiecărui studiu, dintr-o perspectivă generală, sintetică. O recomandare recentă, în cazul existenţei unor multiple ipoteze	Pentru a investiga succesul tehnicii implicite SMP în a surprinde trăsături de personalitate, am selectat pentru acest studiu cazul concret al conştiinciozităţii. În cadrul unui design corelaţional am pus în legătură conştiinciozitatea măsurată implicit cu nivelul explicit al conştiinciozităţii, cu o serie de criterii comportamentale relevante pentru conştiinciozitate (de exemplu, performanţa şcolară), dar şi cu nivelul dezirabilităţii sociale. Dacă tehnica SMP se va dovedi una robustă din perspectivă psihometrică, ar trebui ca aceasta să reproducă relaţiile teoretice dintre constructe (…), să fie

	(mai ales în studii randomizate) este aceea de a diferenţia între ipotezele principale (fundamentale pentru obiectivul cercetării) şi ipotezele secundare (derivate).	fidelă şi să fie rezistentă la variaţii ale nivelului de dezirabilitate socială. În consecinţă, ipotezele studiului sunt: (a) există o asociere pozitivă între nivelurile explicite şi implicite ale conştiinciozităţii; (b) conştiinciozitatea implicită aduce un plus explicativ semnificativ statistic faţă de cât prezice conştiinciozitatea explicită din variaţia comportamentelor specifice persoanelor conştiincioase; (c) nu există o asociere semnificativă statistic între conştiinciozitate, măsurată implicit, şi nivelul de dezirabilitate socială: (d) există corelaţii pozitive ridicate între nivelul conştiinciozităţii implicite măsurat pe acelaşi lot de participanţi, la un interval de şase luni.

Alte consideraţii generale cu referire la partea introductivă:

- atunci când sunt disponibile mai multe referinţe bibliografice pentru a susţine aceeaşi afirmaţie (de exemplu, că există o asociere pozitivă între conştiinciozitate şi performanţa şcolară) se recomandă utilizarea a maximum trei referinţe bibliografice, alese dintre studiile cu cel mai înalt nivel de generalizare/calitativ (de exemplu, metaanalize, studii corelaţionale efectuate pe populaţii mari etc.);
- dacă articolul face parte dintr-un set de studii realizate pe aceeaşi bază de date (de exemplu, un studiu privind eficienţa intervenţiei, un altul despre mecanismele schimbării implicate şi un altul despre raportul cost-eficienţă), este important să se precizeze prin ce anume diferă (ce aduce nou în analiză acest studiu) de cele anterioare;
- dacă articolul vizează un studiu metaanalitic, în partea introductivă ar trebui prezentate informaţii precum: definirea clară a problemei

analizate; un scurt istoric asupra problemei; relevanţa teoretică/aplicativă a metaanalizei; prezentarea şi justificarea selectării variabilelor principale şi a variabilelor moderatoare; discutarea tipurilor de studii găsite din perspectiva designului şi a calităţii acestora; populaţia vizată de metaanaliză; ipotezele studiului[13];

- pentru articolele teoretice (cum ar fi sintezele narative) este greu de decelat o structură standard pentru partea introductivă, deoarece pot exista o varietate de obiective, cum ar fi compararea a două teorii concurente pentru a vedea care este sprijinită mai bine de datele din realitate (sinteză teoretică) sau prezentarea stării de fapt într-un domeniu în scopul stabilirii unor direcţii viitoare de studiu (sinteză a rezultatelor anterioare);
- la obiective (ceea ce se urmăreşte prin studiu) se poate ajunge atât pe bază empirică (pornind de la limitele studiilor anterioare), cât şi pe cale raţională (în baza unor inferenţe/predicţii realizate pornind de la anumite teorii pentru a verifica nivelul de suport al acestora); obiectivele constituie nodul central al unei cercetări, în funcţie de acestea fiind alese designul, participanţii şi variabilele incluse în studiu.

8.1.5. Metodologia cercetării

A doua secţiune majoră dintr-un articol empiric bazat pe modelul IMRAD urmăreşte descrierea metodologiei cercetării (*method*). Deşi această secţiune are un nivel ridicat de structurare – există componente specifice, precum *participanţi*, *instrumente* sau *procedură*, metodologia cercetării este secţiunea în care există cele mai multe cerinţe tipice specifice de raportare, care variază în funcţie de tipul de studiu (experimental,

cvasiexperimental, non-experimental longitudinal, non-experimental comparativ, metaanaliză, evaluări cost-eficienţă etc.). În acest sens au fost dezvoltate o serie de standarde specifice de raportare a rezultatelor în articole de specialitate, cele mai cunoscute fiind prezentate în tabelul 8.2. Trebuie precizat însă că standardele nu fac referire exclusivă doar la secţiunea de metodologie a cercetării, însă aceasta, împreună cu secţiunea de rezultate diferă cel mai mult în funcţie de tipul studiului efectuat.

Tabelul 8.2. *Lista principalelor standarde de raportare existente la ora actuală*

Tip de studiu vizat	Denumirea şi descrierea standardelor	Informaţii de acces/Autori
Studii empirice (recomandări generale)[14]	*JARS* (acronim pentru *Journal Articles Reporting Standards*) a fost dezvoltat de Asociaţia Psihologilor Americani (APA). Modelul se bazează pe o serie de standarde specifice din domenii variate, precum medicină (CONSORT, respectiv TREND) sau educaţie (standardele AERA). JARS este construit modular, fiind la ora actuală format din patru module: tabelul A1.1 (recomandări valabile pentru orice tip de studiu empiric), tabelele A1.2 şi A1.3A pentru studii experimentale, respectiv tabelele A1.2 şi A1.3B pentru studii cvasiexperimentale). Standardele *AERA* (American Educational Research Association) pentru redactarea studiilor empirice din domeniul ştiinţelor sociale au un caracter general, incluzând recomandări atât pentru studii cantitative, cât şi pentru cele care folosesc o metodologie calitativă.	www.apastyle.org/manual/related/JARS-MARS.pdf APA Publications and Communications Board Working Group (2008); Cooper (2011) http://www.aera.net/publications.htm AERA (2006)
Studii randomizate	*CONSORT* (*Consolidating Standards of*	www.consort-statement.org

(experimentale) (capitolul 4)	*Reporting Trials*) este probabil cel mai renumit standard de raportare a rezultatelor construit special pentru redactarea articolelor ce conţin studii experimentale (studii clinice controlate). Deşi standardul a fost dezvoltat de specialişti în medicină, el a fost preluat de multe publicaţii din domeniul psihologiei, fiind la ora actuală cel mai întâlnit model de raportare pentru studii experimentale, fie ele cauzal-descriptive, pragmatice sau orientate spre identificarea mecanismelor cauzale. Pe lângă modelul clasic (actualizat în 2010) există variaţii ale standardelor CONSORT deosebit de utile în psihologie: cazul studiilor de evaluare a intervenţiilor non-farmaceutice, cazul studiilor de evaluare a echivalenţei/non-inferiorităţii etc. (mai multe detalii în capitolul 4).	Modelul clasic (Schulz *et al.*, 2010; Moher *et al.*, 2010); Extensie – randomizare la nivel de cluster (Campbell, 2004); Extensie – studii pragmatice (Zwarenstein *et al.*, 2008) Extensie – studii non-inferioritate/echivalenţă (Piaggio *et al.*, 2006) Extensie pentru intervenţii non-farmaceutice (Boutron *et. al*, 2008a; 2008b).
Studii cvasiexperimentale (capitolul 4)	*TREND* (acronim pentru *Transparent Reporting of Evaluations with Nonrandomized Design*), în principiu, este o adaptare a standardelor CONSORT pentru cazurile în care participanţii nu sunt distribuiţi randomizat în grupele de studiu (grupul experimental, respectiv cel de control), deşi există o manipulare a variabilei independente.	http://www.cdc.gov/trendstatement/ Des Jarlais *et al.* (2004)
Studii non-experimentale (capitolul 5)	*STROBE* (acronim pentru *Strengthening the Reporting of Observational Studies in Epidemiology*) este un standard dezvoltat în domeniul ştiinţelor medicale (actualizat în 2007) care poate fi utilizat ca ghid pentru redactarea articolelor empirice bazate pe un design	www.strobe-statement.org von Elm *et al.* (2007); Vandenbroucke *et al.* (2007)

	de tip non-experimental. Sunt luate în considerare aspecte specifice pentru trei tipuri de studii: longitudinale, studii transversale corelaţionale, studii transversale ce vizează comparaţii *ex post facto* (*case-control studies*).	
Studii de analiză a eficienţei economice (capitolul 6)	*RCT-CEA* (acronim pentru *Randomized Clinical Trial – Cost Effectiveness Analysis*) este unul dintre standardele dezvoltate pentru redactarea articolelor care urmăresc analize de tip cost-eficienţă sau cost-utilitate rezultate în baza unor studii experimentale, preferabil de tip pragmatic.	http://www.ispor.org/workpaper/clinical_trial.asp Ramsey *et al.* (2005)
Studii metaanalitice[12] (capitolul 7)	*PRISMA* (acronim pentru *Preferred Reporting Items for Systematic Reviews and Metaanalyses*) (variantă îmbunătăţită a QUOROM) reprezintă cel mai complex standard de raportare (redactare) a studiilor metaanalitice, cu accent pe metaanaliza studiilor de tip experimental şi cvasiexperimental. *MOOSE* (acronim pentru *Metaanalysis of observational studies in epidemiology*) conţine standardul de raportare adaptat pentru metaanalize efectuate pe studii de tip non-experimental. *MARS* (acronim pentru *Metaanalysis Reporting Standards*) a fost dezvoltat sub girul Asociaţiei Psihologilor Americani (APA), modelul fiind inspirat din prevederile PRISMA şi MOOSE. Modelul reprezintă la ora actuală varianta recomandată de APA pentru raportarea metaanalizelor în psihologie. Are însă un caracter mai general, fără a diferenţia între metaanalizele bazate pe	www.prisma-statement.org Moher *et al.* (2009) Stroup *et al.* (2000) www.apastyle.org/manual/related/JARS-MARS.pdf APA Publications and Communications Board Working Group (2008); Cooper (2011)

	studii experimentale şi cele bazate pe studii non-experimentale.

Prin urmare, este extrem de dificil să se ofere un set de recomandări care să fie general valabile pentru orice tip de studiu, secţiunea de metodologie şi cea de rezultate depinzând cel mai mult de specificul fiecărui tip de studiu. De aceea, prima recomandare către cititori este de a consulta direct standardele din tabelul 8.2 corespunzătoare tipului de studiu pe care îl au de redactat.

Totuşi, în scop ilustrativ, voi prezenta câteva linii directoare de inclus în secţiunea metodologică a unui articol empiric, pornind de la principiile comune specificate în primul tabel (A1.1) al standardului JARS (Cooper, 2011), dar fără a adera strict la toate detaliile cuprinse în acesta, pentru a permite focalizarea pe aspectele esenţiale de inclus/abordat în această secţiune a unui articol empiric.

Pe ansamblu, calitatea principală pe care trebuie să o aibă secţiunea metodologică a unui articol empiric este de a prezenta într-un mod explicit activităţile şi resursele utilizate în cercetarea respectivă pentru a obţine rezultatele prezentate. Dacă se constată deficienţe la nivelul concepţiei şi implementării studiului (aspecte metodologice, tehnice), atunci este foarte probabil ca rezultatele să fie distorsionate. De asemenea, este important să se observe la acest nivel în ce măsură obiectivele studiului pot fi atinse cu ajutorul metodologiei utilizate. Tocmai din aceste motive, pentru a aduce un plus de claritate în ochii cititorilor, secţiunea metodologică este adesea împărţită adesea în trei-patru secţiuni distincte.

Astfel, am putea considera că rolul principal al secţiunii metodologice într-un articol empiric este de a: (1) descrie resursele utilizate (date despre sursa informaţiilor şi despre metodele utilizate pentru colectarea lor –

participanţi, respectiv *instrumente*); (2) descrie demersul de colectare şi analiză a datelor (*procedură*[15] şi uneori *design*[16]).

Tabelul 8.3. *Descrierea celor trei componente tipice ale secţiunii de metodologia cercetării (adaptare selectivă a datelor prezentate în primul studiu din Sava* et al. *(2012)*

Subsecţiune	Rol şi recomandări	Exemplificare (prezentare selectivă/adaptată a unor fragmente de text)
Participanţi		
Criterii de includere/excludere	Îndeosebi în studiile experimentale sunt utilizate explicit o serie de criterii de includere, respectiv excludere a indivizilor din studiu. Dacă există asemenea criterii, ele trebuie precizate explicit. Rolul precizării unor asemenea criterii este relevant pentru a controla anumite variabile care ar putea afecta rezultatele studiului (de exemplu, într-un studiu privind impactul ADHD asupra performanţei şcolare sunt incluşi elevii diagnosticaţi cu ADHD, cu condiţia să aibă un IQ > 85, în baza unei testări iniţiale).	Lotul de participanţi a fost selectat dintr-un eşantion iniţial format din 120 de studenţi supuşi unei evaluări psihologice. Criteriile de eligibilitate pentru studiu au fost următoarele: (a) participanţii să se afle la prima lor experienţă de evaluare printr-o măsură de tip implicit; (b) să nu aibă în istoricul personal vreo suferinţă care să fi necesitat intervenţie psihiatrică; (c) să nu prezinte vreo tulburare de personalitate din axa II a DSM-IV, în baza evaluării psihologice realizate. De asemenea, pentru a fi incluşi în analiză participanţii trebuiau să completeze sarcina SMP de evaluare a conştiinciozităţii la nivel implicit, conform instructajului, fiind excluşi cei care au completat superficial sarcina (care au avut la mai mult de 10% din cazuri un timp de reacţie pentru răspuns sub 100 ms).
Caracteristicile participanţilor	Aici sunt prezentate informaţii utile despre caracteristicile participanţilor incluşi în studiu: de tip demografic (vârstă, sex, status [lot clinic *vs* lot neclinic]) sau	Astfel, în final au fost analizate rezultatele provenite de la 98 de participanţi (82%). Cinci studenţi au fost excluşi din analiză pentru că nu îndeplineau criteriile de eligibilitate, iar

	variabile relevante pentru problema studiată (dacă este cazul).	alţi 17 pentru nerespectarea instructajului. Caracteristicile demografice pentru cei 98 de participanţi incluşi în analiză indică o proporţie mai ridicată de studenţi de sex feminin (75%) şi vârste cuprinse între 19 şi 31 de ani (M = 21,2; SD = 3,8).
Descrierea procedurii de eşantionare	Se precizează ce fel de procedură a fost utilizată pentru constituirea eşantionului de participanţi.	Procedura de recrutare a studenţilor a fost una de tip neprobabilistic, bazată pe convenienţă, în baza unui anunţ postat electronic pe site-ul universităţii.
Precizări referitoare la aspectele etice	Se referă la consimţământul informat de a fi subiecţi ai cercetării şi la eventualele beneficii oferite.	Participanţii eligibili pentru acest studiu au semnat un formular de consimţire informată a participării lor la studiu şi au fost recompensaţi cu un punctaj pentru efortul lor voluntar.
Alte precizări tehnice	O ultimă componentă de inclus cu ocazia descrierii participanţilor face referire la puterea statistică a studiului. Tot aici trebuie precizat dacă există alte publicaţii pornind de la eşantionul studiat (dacă este cazul).	Deoarece este prima dată când se încearcă utilizarea sarcinii SMP pentru a măsura implicit nivelul conştiinciozităţii, se observă că cei 98 de participanţi analizaţi asigură o putere statistică de 0,70 pentru a surprinde asocieri de intensitate medie (moderată, r = 0,25), între această variabilă centrală şi măsura explicită a conştiinciozităţii/criteriile comportamentale incluse.
Instrumente (măsurători) – Ilustrare selectivă		
	Fie că este vorba de teste psihologice, de sarcini creative, de măsurători fiziologice etc., sunt prezentate succesiv, pornind de la cele mai relevante, cum ar fi conceptul central al lucrării sau variabilele dependente principale (*primary outcomes*), şi finalizând	Sarcina SMP a fost utilizată pentru a surprinde nivelul implicit al conştiinciozităţii (C SMP). Pentru a implementa C SMP am apelat la aplicaţia software Direct RT v 2008 (Empirisoft, 2008). Sarcina implică prezentarea succesivă pe ecranul calculatorului a 32 de

	cu cele cu o relevanţă mai scăzută. În cazul testelor, modelul clasic de redactare implică o trimitere bibliografică către instrumentul utilizat, scurte detalii legate de interpretare şi referiri succinte la caracteristicile psihometrice. Totuşi, dacă probele sunt mai puţin cunoscute, acest ultim aspect poate fi dezvoltat în cadrul unui paragraf întreg. În cazul utilizării unor aparaturi sau în situaţia apelării la unele sarcini de laborator, acestea sunt descrise în detaliu, dacă nu există vreo referinţă bibliografică ce conţine aceste detalii.	perechi de stimuli (amorsă + ţintă). Stimulii-amorsă fac referire la 16 descriptori adjectivali specifici unui nivel ridicat al conştiinciozităţii (de exemplu, sârguincios), şi la 16 descriptori adjectivali relevanţi pentru un nivel scăzut al conştiinciozităţii (de exemplu, leneş). Aceştia au fost selectaţi dintre descriptorii adjectivali recunoscuţi ca specifici pentru cei doi poli ai C (Schmuckle şi Egloff, 2008). Stimulii-ţintă constau într-o serie de pictograme chinezeşti, neutre din punct de vedere semantic (datorită opacităţii lor) şi afectiv. În acest sens, toate pictogramele chinezeşti au fost evaluate ca fiind neutre afectiv într-un pretest în care li se cerea participanţilor să evalueze pe o scală de la 1 (neplăcut) la 7 (plăcut) valenţa fiecărei pictograme în parte (N = 50, M = 4,26, SD = 0,31).
	De asemenea, sunt foarte utile precizările suplimentare cu privire la calitatea datelor colectate, îndeosebi pentru sarcini mai puţin standardizate sau pentru cele construite ad-hoc. Ne referim aici, fără a ne rezuma doar la acestea, la aspecte precum: dovezi din pretestare cu referire la anumite proprietăţi ale instrumentelor; la experienţa celor care implementează intervenţiile psihoterapeutice; performanţa în evaluarea/codarea observaţiilor în baza unor grile; verificarea reuşitei manipulării experimentale.	Cele 32 de amorse au fost prezentate într-o ordine aleatorie, fiind împerecheate, tot aleatoriu, cu stimulii-ţintă. În fiecare dintre cele 32 de secvenţe, stimulul-amorsă a fost prezentat în centrul ecranului, cu caractere Arial 16, pentru o perioadă de 200 ms. Stimulul era apoi înlocuit cu un dreptunghi alb (5 × 2 cm) timp de 125 ms, totalizând un SOA de 325 ms. După dispariţia dreptunghiului alb, apare stimulul-ţintă timp de 200 ms, cu dimensiunea pictogramei de 5 × 2 cm. Apoi pictograma este înlocuită de un dreptunghi de culoare gri de aceleaşi dimensiuni, care rămâne pe ecran până la înregistrarea răspunsului.

		Aspectul inovativ implicat de SMP este faptul că li se cere participanţilor să evalueze pentru fiecare pictogramă în parte dacă aceasta ar fi potrivită să fie imprimată pe un tricou personalizat sau nu. În evaluarea dacă pictograma i se potriveşte sau nu, i se atrage atenţia celui testat să nu se lase influenţat în decizia sa de semnificaţia stimulului-amorsă care precedă fiecare pictogramă evaluată. Pentru a cota C SMP s-a acordat câte un punct pentru fiecare pictogramă aleasă ca fiind potrivită, dacă aceasta era precedată de un stimul-amorsă din zona polului ridicat al C, respectiv pentru fiecare pictogramă respinsă (nepotrivită), dacă aceasta era precedată de un stimul-amorsă din zona polului scăzut al C. Astfel, scoruri ridicate sunt indicative pentru un nivel crescut al conştiinciozităţii măsurate implicit. Consistenţa internă măsurată în acest caz prin metoda înjumătăţirii indică o valoare acceptabilă pentru C implicit, de 0,64). Scala C din proba NEO-FFI (Costa şi McRae, 1992) a fost utilizată pentru surprinderea nivelului explicit al conştiinciozităţii. Scala C conţine 12 itemi în format Lickert cu 5 trepte, scorul total fiind obţinut prin însumarea celor 12 itemi. Scoruri ridicate la scala C din NEO-FFI indică un nivel crescut al C explicite, consistenţa internă pentru acest studiu fiind de 0,76.
Procedură		
Aspecte ce ţin de	Sunt oferite informaţii cu privire la	Participanţii au fost testaţi în varianta

colectarea datelor	contextul cercetării (locaţie, contextul testării, succesiunea probelor), iar uneori detalii suplimentare (cum ar fi instructajul verbatim) pentru măsurători esenţiale sau o referinţă bibliografică pentru acestea.	creion-hârtie, în grupuri de 5-10 persoane, în cazul măsurilor explicite, respectiv individual, la calculator, într-o cameră izolată fonic, în cazul C SMP. Ordinea de prezentare a instrumentelor a fost: mai întâi C SMP, apoi setul de probe explicite (acestea din urmă fiind contrabalansate ca ordine de prezentare).
Aspecte ce ţin de analiza datelor	Se referă la managementul datelor colectate (date care lipsesc, prelucrări preliminare ale lor etc.)	Pentru a analiza sintetic relaţia dintre C SMP, respectiv C NEO-FFI, cu indicatorii comportamentali, aceştia au fost regrupaţi prin metoda componentelor principale într-un singur scor global.

Dintr-o perspectivă sintetică, această secţiune metodologică ar trebui să ne ajute să vedem esenţa calităţii studiului, cel puţin cu privire la: (a) „Pe ce populaţie a fost făcut studiul şi în ce măsură rezultatele pot fi generalizate în alte contexte?” (validitate externă); (b) „Ce fel de măsurători au fost efectuate şi ce încredere avem în aceste măsurători, respectiv în procedurile de colectare a lor?” (validitate de construct); (c) „În ce măsură designul ales şi variabilele incluse sunt potrivite pentru a răspunde adecvat la obiectivul studiului?” (validitate internă); (d) „În ce măsură informaţiile prezentate în această secţiune sunt suficient de detaliate pentru a permite replicarea studiului (validitate externă) şi pentru a aprecia ca reduse riscurile obţinerii unor rezultate distorsionate, ca urmare a unor deficienţe în implementarea studiului şi/sau a necontrolării unor variabile străine (validitate internă) sau a unui număr insuficient de participanţi (validitate statistică)?”[17].

Dată fiind importanţa deosebită a acestor întrebări, este de la sine înţeles că modelul prezentat în tabelul 8.3 este un model minimal, util mai

degrabă pentru a descrie studii non-experimentale de tip corelaţional. Aceste specificaţii trebuie adaptate în funcţie de caracteristicile fiecărui tip de studiu în conformitate cu prevederilor specificate în standardele asociate acestora (vezi tabelul 8.2).

8.1.6. Rezultate

A treia secţiune majoră dintr-un articol empiric este aceea în care sunt prezentate rezultatele. La fel ca în cazul secţiunii de metodologie a cercetării, structura secţiunii de rezultate variază mult în funcţie de tipul de studiu derulat şi de problematica investigată (obiectivul urmărit).

De aceea, se recomandă utilizarea aceloraşi standarde specifice specificate în tabelul 8.2 pentru a redacta acurat această parte din articol. În cele ce urmează vom prezenta câteva sugestii generale cu privire la redactarea acestei secţiuni, dar şi unele aspecte punctuale care ar trebui să se regăsească în secţiunea de rezultate.

Menţiuni generale pentru secţiunea de rezultate (Priebe, 2008; Thompson, 1999):

- de regulă, secţiunea de rezultate conţine: (a) paragrafe de text în care se expune semnificaţia datelor prezentate, însoţită de dovezi statistice (tehnice) care să o susţină (de exemplu, *se constată un nivel mai scăzut al anxietăţii sociale la participanţii care au urmat terapie cognitiv-comportamentală decât la cei din grupul de control – $t(23) = 3{,}43$, $p < 0{,}01$, $d = 1{,}37$*); (b) tabel sau tabele sumare cu rezultatele studiului[18]; (c) figură sau figuri ilustrative cu cele mai importante rezultate ale studiului (de exemplu, cu efectul observat asupra variabilei dependente principale) (opţional); din acest punct de vedere, informaţiile numerice din tabele sau grafice nu trebuie trecute şi în format text[19], pentru a evita redundanţa[20];

- rolul acestei secţiuni este de a prezenta într-un mod obiectiv (fără a interpreta sau a specula asupra rezultatelor studiului, această parte revenind secţiunii de discuţii) şi nedistorsionat rezultatele obţinute (ar trebui prezentate toate rezultatele studiului, atât rezultatele care sprijină, cât şi cele care nu sprijină ipotezele de lucru);
- deşi multe studii realizate dispun de o colecţie impresionantă de rezultate, este important ca accentul să cadă pe prezentarea aspectelor esenţiale pentru obiectivul vizat; analizele secundare pot fi discutate succint sau pot fi prezentate în tabele[21], chiar fără a fi discutate dacă există un spaţiu editorial restrâns;
- cu cât obiectivele studiului sunt mai îndrăzneţe (identificarea unei relaţii de tip cauză-efect, relevarea unor mecanisme ale schimbării), cu atât secţiunea de rezultate este mai extinsă, fiind analizate nu numai aspectele legate de întrebarea esenţială („Este intervenţia A mai bună decât intervenţia B?”), ci şi cele care ar putea conduce la răspunsuri distorsionate la această întrebare (procentul de renunţare la experiment; aderenţa la protocolul de intervenţie; complianţa participanţilor etc.; pentru detalii, vezi standardele CONSORT în Moher *et al.*, 2010); cu alte cuvinte, pe lângă rezultatele obţinute cu referire la problema centrală de investigat, ar trebui analizate şi prezentate orice alte date care ne oferă indicii cu privire la posibile probleme privind validitatea rezultatelor obţinute (fie ea internă, externă, statistică sau de construct);
- din perspectivă stilistică, ar fi de dorit ca fiecare element important analizat să fie prezentat distinct, într-un paragraf separat, eventual însoţit de o etichetă lingvistică care să denumească aspectul vizat (dacă este cazul).

Tabelul 8.4. *Descrierea componentelor tipice ale secţiunii de rezultate. Adaptare selectivă a rezultatelor (sunt excluse tabelele) după Sava* et al. *(2012)*

Subsecţiune	Rol şi recomandări	Exemplificare (prezentare selectivă/adaptată a unor fragmente de text)
Reiterarea obiectivului principal al studiului şi a analizelor statistice necesare	Secţiunea de rezultate poate începe cu reamintirea punctului central de investigat şi a analizelor necesare de efectuat, luând în calcul obiectivele şi metodologia studiului. Suplimentar, în contextul apelării la tehnici statistice mai puţin cunoscute, se poate introduce o frază care să specifice (justificarea utilizării) tehnica respectivă şi elementele-cheie utile pentru interpretare.	Proprietăţile psihometrice (validitate de construct, validitate de criteriu şi fidelitate) ale SMP au fost evaluate asociind măsura implicită a conştiinciozităţii (C) cu măsuri explicite ale acesteia, cu o serie de comportamente specifice C şi cu nivelul dezirabilităţii sociale, cu ajutorul tehnicilor de corelaţie simplă Bravais-Pearson şi a regresiei multiliniare ierarhice, prelucrarea datelor fiind realizată în SPSS versiunea 17.0.
Analizele preliminare	Scopul paragrafului este de a prezenta cititorilor eventuale analize preliminare necesare (de exemplu, construirea unui indicator global), precum şi analize informative din perspectiva validităţii rezultatelor (de exemplu, oferirea dovezilor cu privire la echivalenţa unor grupe comparate la nivel de vârstă, gen sau alte variabile străine relevante/decizii de eliminare a unor cazuri/variabile din analiză şi justificarea deciziei); managementul datelor care lipsesc.	Deoarece atât pentru surprinderea lui C măsurată explicit, cât şi a comportamentelor asociate acestei trăsături au fost evaluaţi mai mulţi indicatori manifeşti, pentru simplificarea analizei am apelat la analiza factorială exploratorie, metoda componentelor principale, pentru a obţine câte un indicator global pentru C explicit, respectiv pentru C comportamente specifice. Ambele scoruri obţinute reprezintă scoruri standardizate, având media egală cu 0 şi abaterea standard egală cu 1. Vom preciza însă că doar 4 dintre cele 6 comportamente selectate au fost incluse în indicatorul global selectat, două comportamente fiind eliminate din analiză din cauza variabilităţii scăzute a rezultatelor (de exemplu, 98%, respectiv 96% dintre participanţii au raportat că au chiulit cel puţin o dată de la şcoală în ultima lună, respectiv că nu au consumat droguri ilegale în ultimele două săptămâni).

Discutarea succesivă a ipotezelor, pornind de la cea principală	Fiecare ipoteză formulată este discutată într-un paragraf distinct, prezentându-se toate informaţiile necesare legate de acestea. În cazul în care datele obţinute pentru o ipoteză anume sunt contradictorii se prezintă atât datele care susţin ipoteza, cât şi cele care nu constituie un sprijin pentru aceasta. Această precizare este valabilă pentru fiecare ipoteză în parte, după schema: Ipoteza 1 (principală): rezultate (care sprijină ipoteza, care nu susţin ipoteza, alte precizări tehnice, dacă este cazul) Ipoteza 2: rezultate… … Ipoteza n: rezultate…	Datele prezentate în tabelul 1 sprijină validitatea de construct a C SMP. Astfel, C SMP corelează pozitiv, la un nivel semnificativ statistic, atât cu nivelul global al lui C măsurat explicit, cât şi cu comportamentele autoraportate asociate lui C. Mai mult, nivelul implicit de conştiinciozitate (C SMP) are validitate incrementală, aducând un plus explicativ pentru relaţia dintre nivelul explicit de conştiinciozitate şi comportamentele asociate cu C, jucând un rol moderator. Astfel, se constată o asociere mai intensă între C explicit şi comportamentele asociate la participanţii cu niveluri congruente ale lui C (C implicit ridicat şi C explicit ridicat, respectiv C implicit scăzut şi C explicit scăzut; vezi figura 1 şi tabelul 2)[22]. Totuşi, valididitatea incrementală a C SMP este prezentă doar la nivelul efectului de interacţiune (moderator), nu şi la nivelul efectului principal (aditiv). Astfel, dacă excludem efectul de moderare, C SMP nu reuşeşte să explice modul în care variază comportamentele asociate lui C dincolo de ceea ce a fost deja explicat de către C explicit. Un alt aspect analizat a fost relaţia dintre C SMP şi nivelul dezirabilităţii sociale…
Alte rezultate relevante, neincluse iniţial ca ipoteze	În cazul în care se obţin unele rezultate importante care nu erau cuprinse iniţial în vreo ipoteză sau în cazul în care sunt realizate analize secundare care să ofere explicaţii plauzibile pentru anumite rezultate obţinute (de exemplu, analize *post hoc*), fără ca acestea să fi fost specificate iniţial ca ipoteze ale cercetării.	Date interesante rezultă în urma aplicării strategiei analitice de a pune în legătură C implicit, respectiv C explicit cu fiecare dintre cele patru comportamente incluse în indicatorul comportamental global (vezi tabelul 3). Astfel, se constată că atât C explicit, cât şi C implicit corelează direct cu (aceleaşi) trei comportamente autoraportate din cele patru incluse în analiză, aspect ce susţine validitatea convergentă a celor două tipuri distincte de măsuri ale C.

Concluzionând, rolul secţiunii rezultatelor este de a oferi strict răspunsuri obiective, cu privire atât la obiectivul şi la ipotezele formulate („Sunt acestea sprijinite?”), cât şi la alte aspecte relevante (verificarea asumpţiilor asociate unor tehnici statistice; controlul unor variabile străine; managementul datelor care lipsesc etc.) pentru a stabili nivelul de validitate al acestor rezultate („Cât de mult ne putem încrede în aceste rezultate?”), precum şi pentru a colecta dovezi suplimentare care vor facilita interpretarea rezultatelor obţinute – analize *post hoc*. Analiza calitativă a acestor rezultate are loc abia în secţiunea următoare, destinată discuţiilor (în engleză, se preferă utilizarea singularului – *discussion*).

8.1.7. Discuţii

Unii autori (Budgell, 2009; Sava, 2005) consideră această secţiune ca fiind cea mai importantă parte a unui articol empiric, deoarece în această secţiune interpretăm (acordăm un sens) rezultatele obţinute pe baza metodologiei de cercetare aplicate. Dacă partea de metodologie şi rezultate era una eminamente factuală (prezentarea a ceea ce a fost întreprins şi a rezultatelor obţinute, fără consideraţii de ordin calitativ), secţiunea de discuţii este una interpretativă, analizând calitativ metodologia derulată, rezultatele studiului şi implicaţiile acestora din perspectivă teoretică sau practică.

De asemenea, spre deosebire de secţiunile anterioare referitoare la metodologie şi rezultate, partea de discuţii are o structură standardizată, valabilă pentru cele mai multe tipuri de studii, desigur, cu unele particularităţi specifice tipului de studiu realizat (de exemplu, un accent mai mare pe aspecte de validitate internă în studiile care vizează o relaţie cauză-efect).

Tabelul 8.5. *Descrierea părţilor componente tipice din secţiunea de discuţii. Ilustrarea selectivă a acestor aspecte este o adaptare după Sava* et al. *(2012)*

Subsecţiune	Rol şi recomandări	Prezentare selectivă/adaptată a unor paragrafe/fragmente de text în scop ilustrativ
Reiterarea obiectivului major al studiului şi oferirea unui răspuns succint la întrebarea (întrebările) cercetării	Secţiunea de discuţii începe întotdeauna cu precizarea obiectivului urmărit în studiu şi a rezultatelor principale obţinute, cu referire la întrebarea majoră a cercetării. Dacă există şi alte întrebări (obiective secundare), se fac referiri succinte şi la rezultatele obţinute în cazul acestora, însă accentul este pus îndeosebi pe problema principală de investigat.	Scopul articolului a fost de a introduce şi valida o nouă procedură implicită de evaluare a conştiinciozităţii, denumită SMP (*Semantic Misattribution Procedure*), această tehnică fiind derivată din AMP (*Affect Misattribution Procedure*). Datele obţinute susţin validitatea SMP, iar atunci când numărul de secvenţe de răspuns se dublează, SMP dovedeşte şi o bună fidelitate, atât în termeni de consistenţă internă, cât şi de stabilitate test-retest. De asemenea, s-a constatat că tehnica SMP este mai rezistentă, dar nu şi imună, la tendinţa de a oferi răspunsuri dezirabile social, comparativ cu metodele explicite de investigaţie.
Oferirea de explicaţii *post hoc* pentru rezultatele neaşteptate obţinute (dacă este cazul)	O secvenţă cu caracter opţional vizează oferirea unor explicaţii *post hoc* în cazul relevării unor rezultate contrare aşteptărilor sau nesemnificative statistic (nesprijinirea ipotezele). Dacă toate rezultatele converg în direcţia aşteptată nu este nevoie de această discuţie *post hoc*, deoarece logica (motivul) pentru acest rezultat a fost prezentată în partea introductivă, în momentul justificării ipotezelor formulate.	Cu privire la acest ultim aspect, o posibilă explicaţie pentru faptul că tehnica SMP nu este imună la tendinţa de a oferi răspunsuri dezirabile social are în vedere discuţia introdusă de De Houwer (2006). Acesta preciza că nu se poate vorbi de tehnici implicite pure, ele fiind situate mai degrabă pe un continuum, în funcţie de caracteristica vizată (de exemplu, controlabilitate, conştientizare etc.), de la implicit spre explicit. O altă explicaţie posibilă este…
Integrarea rezultatelor din acest	O secvenţă importantă şi obligatorie vizează compararea rezultatelor obţinute cu cele din alte studii focalizate pe	Rezultatele obţinute în acest studiu converg cu cele obţinute de Schmukle *et al.* (2008), relevând că măsurile implicite

studiu în literatura de specialitate	aceeaşi temă. Sunt amintite atât studiile care obţin rezultate convergente, cât şi cele cu rezultate divergente. În cazul unor rezultate contradictorii se pot oferi explicaţii (speculative) cu privire la cauzele acestora.	ale personalităţii au o validitate predictivă incrementală, reuşind să prezică anumite tendinţe comportamentale dincolo de ceea ce era deja prezis prin măsurile explicite ale personalităţii… Spre deosebire de alte studii (Asendorpf, 2002; Banse *et al.*, 2006), în studiul nostru au fost obţinute corelaţii semnificative între măsurile explicite şi cele implicite ale lui C, acest rezultat fiind replicat în toate cele trei studii incluse, sugerând astfel că între măsurile explicite şi cele implicite ale unei trăsături de personalitate nu există o relaţie de independenţă…
Valoarea adăugată a studiului	Pe baza discutării relaţiei studiului cu alte cercetări similare este foarte important să se precizeze, conform cu rezultatele obţinute, ce aduce nou acest studiu, care ar fi punctele sale tari, aspect important din perspectiva deciziei unui editor de a publica sau nu studiul. Valoarea adăugată poate viza aspecte teoretice, metodologice sau practice.	Spre deosebire de studiile anterioare centrate pe măsurători implicite din sfera testului asocierilor implicite (IAT) (Schmuckle *et al.*, 2008) şi a derivatelor acestuia (GNAT) (Boldero, 2007), acest studiu reprezintă prima încercare de a evalua implicit nivelul conştiinciozităţii prin intermediul SMP, o tehnică ce funcţionează pe baza unui alt mecanism cognitiv (atribuirea greşită) decât cele bazate pe puterea asocierilor. Demersul pare a fi unul încununat cu succes, tehnica SMP fiind superioară tehnicilor derivate din IAT şi cel puţin similară cu cea a IAT, dacă luăm în considerare criteriul calităţii proprietăţilor psihometrice…
Implicaţii teoretice şi/sau practice ale rezultatelor	De regulă, acestea ar trebui să aibă în vedere elementele prezentate în secţiunea introductivă, unde s-a expus relevanţa problemei investigate.	Din perspectivă teoretică, rezultatele obţinute ridică semne de întrebare cu privire la relevanţa modelor interpretative din domeniul atitudinilor (modelul dual *vs* singular) în interpretarea datelor

		rezultate din aplicarea măsurilor implicite la nivelul trăsăturilor de personalitate, deoarece niciunul dintre cele două modele nu reuşeşte să explice decât parţial rezultatele obţinute…
Limitele studiului	Deoarece orice cercetare are limitele sale, acestea ar trebui precizate pentru a identifica riscul obţinerii unor concluzii distorsionate. Aceste limite pot viza aspecte diverse, în funcţie de tipul de studiu, printre cele mai întâlnite fiind: (a) nivelul validităţii interne (de exemplu, un număr ridicat de persoane renunţă să participe până la final în intervenţie etc.); (b) nivelul validităţii de construct (utilizarea unui instrument necunoscut etc.); (c) nivelul validităţii statistice (rezultate inconsistente, putere statistică redusă etc.); (d) nivelul validităţii externe (generalizare rezultate etc.). Acolo unde este posibil, limitele amintite pot fi dezamorsate, dacă există dovezi obiective în această direcţie (vezi detalii despre riscurile privitoare la validitatea unui studiu în capitolul 3).	Deoarece în sarcina SMP li se solicită participanţilor să evalueze doar stimulii-ţintă (pictogramele), este posibil ca o parte dintre participanţi să ignore acest instructaj şi să evalueze stimulii-amorsă (adjectivele), acestea având un conţinut semantic clar, spre deosebire de caracterul ambiguu al pictogramelor. Dacă acest lucru apare, atunci SMP nu mai poate fi catalogată ca fiind o tehnică implicită, ci una explicită, de evaluare directă. Totuşi, luând în considerare că nivelul corelaţiilor dintre SMP C şi C explicit în cele trei studii este situat în jurul valorii de 0,23 (valoare similară întâlnită în cazul altor tehnici implicite care nu prezintă riscul de a fi evaluate direct), este puţin probabil ca acest aspect să fie un risc puternic la nivelul concluziilor studiului… O altă limită are în vedere contextul de laborator în care s-a derulat studiul. Astfel, este posibil ca dinamica implicit-explicit şi relaţia dintre aceasta şi dezirabilitatea socială să fie alta într-un context ecologic, cum ar fi cel al testării ca parte a selecţiei de personal, din cauza consecinţelor (mizei) diferite pentru participanţi…
Propunerea unor studii viitoare	Pornind de la limitele studiului şi luând în calcul statusul obiectivului investigat din perspectivă teoretică, metodologică sau practică, pot fi specificate câteva	Studiile viitoare pe această temă pot urmări impactul pe care lungimea probei SMP, operaţionalizată ca număr de secvenţe, îl are asupra proprietăţilor sale

	direcţii viitoare de studiu (obiective ce decurg din studiul derulat sau care sunt necesare pentru a elimina unele neajunsuri ale studiului realizat).	psihometrice. Studiul de faţă sugerează cum dublarea numărului de secvenţe poate contribui la augmentarea consistenţei interne a probei şi la intensificarea corelaţiei dintre măsurile implicite şi cele explicite de personalitate. Replicarea acestor rezultate, prin studii viitoare, este necesară pentru a dezvolta în timp un instrument psihometric cu validitate crescută, dar care să rămână la fel de rezistent la înclinaţia unor persoane de a oferi răspunsuri dezirabile social…
Mesaj concluziv	Se recomandă ca articolul să se încheie cu un scurt paragraf sintetic, al cărui rol este similar cu al celui existent în rezumatul articolului. Parafrazând o expresie din limba engleză, acesta ar fi *the take-away message* pe care cititorul ar trebui să-l reţină în urma lecturării articolului.	Concluzionând, în ciuda unor aspecte metodologice care trebuie rezolvate pentru a permite implementarea acestui gen de măsurători în practica curentă a evaluării psihologice, SMP se anunţă ca o tehnică implicită promiţătoare pentru studiul personalităţii, dovedind calităţi psihometrice similare cu cele ale IAT, respectiv mai bune decât cele observate în cazul celorlalte tehnici asociative derivate.

Alte menţiuni generale pentru secţiunea de discuţii:

- date fiind atât caracterul interpretativ al acestei secţiuni (ce semnificaţie teoretică şi aplicativă pot avea rezultatele obţinute?), cât şi numeroasele limite care pot afecta validitatea concluziilor formulate, secţiunea de discuţii apelează frecvent la expresii lingvistice cu caracter probabilistic/prudente, precum „datele sugerează faptul că”, „o explicaţie posibilă ar fi”, „se pare că”; „rezultatele ridică semne de întrebare cu privire la afirmaţia anterioară…” etc.; această practică permite formularea unor explicaţii

cu caracter speculativ care să faciliteze integrarea rezultatelor obţinute în cadrul teoretic general relevant pentru problema investigată;

- cu excepţia primului şi ultimului aspect menţionat (reiterarea obiectivului studiului, respectiv mesajul concluziv), celelalte componente pot avea altă ordine sau pot fi chiar intercalate, cu condiţia ca aspectele de discutat să rămână aceleaşi (de exemplu, ipotezele formulate pot fi discutate pe rând, prezentându-se implicaţiile teoretice şi/sau practice, limitele aferente şi sugestiile viitoare de studiu pentru fiecare ipoteză analizată);
- calitatea discuţiilor aduce adesea un plus semnificativ unui articol prin capacitatea autorului de a releva semnificaţii teoretice/practice pentru rezultatele obţinute[23]; de aceea funcţia esenţială a discuţiilor este de a integra rezultatele obţinute în direcţia obţinerii unui progres în cunoaştere, justificând astfel importanţa (utilitatea) studiului respectiv din punct de vedere teoretic, metodologic sau aplicativ.

8.1.8. Bibliografie

De cele mai multe ori, atunci când nu există şi anexe, bibliografia este ultima secţiune a unui studiu de specialitate. Standardele de redactare recunoscute internaţional în domeniul psihologiei sunt cele elaborate de Asociaţia Psihologilor Americani, ajunse în prezent la ediţia a VI-a (APA, 2010). Aceste standarde nu vizează doar componenta bibliografică, ci au în vedere atât aspecte care ţin de modalitatea de citare („Cum se citează?” şi „Cât se citează?”; vezi secţiunea 8.2.2), cât şi alte chestiuni importante pentru autorii articolelor de specialitate.

O simplă activitate de căutare pe internet va evidenţia mai multe materiale educative disponibile gratuit despre cum se scrie secţiunea de

bibliografie conform standardelor APA; de aceea, nu vom insista aici asupra acestui aspect, ci vom prezenta succint trei dintre cele mai întâlnite tipuri de referinţe bibliografice[24].

Tabelul 8.6. *Trei tipuri de referinţe bibliografice des întâlnite, redactate conform celei de-a şasea ediţii a standardelor APA (2010)*

Tip sursă	Exemplificare
Articol dintr-o revistă de specialitate[25]	Sava, F.A., & Sperneac, A.M. (2006). Sensitivity to reward and sensitivity to punishment rating scales: A validation study on the Romanian population. *Personality and Individual Differences*, *41*, 1445-1456. doi:10.1016/j.paid.2006.04.024
Capitol dintr-o carte de specialitate	Szentagotai, A., & Jones J. (2010). The behavioral consequences of irrational beliefs. In D. David, S.J. Lynn, & A. Ellis (Eds.), *Rational and irrational beliefs. Research, theory, and clinical practice* (pp. 75-98). New York, NY: Oxford University Press.
Carte de specialitate	Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to MetaAnalysis*. Chichester, UK: Wiley.

Pe lângă consultarea manualului (APA, 2010) şi a materialelor informative cu privire la acest subiect disponibile gratuit pe internet, o resursă deosebit de utilă în această direcţie este aplicaţia gratuită Mendeley[26] (Mendeley, s.a.). Una dintre caracteristicile de bază ale acestui program este aceea de a permite utilizatorilor să genereze automat bibliografia unui articol pe baza citărilor introduse în text, inclusiv în formatul cerut de APA. Mai mult decât atât, aplicaţia permite importarea informaţiilor (detaliilor bibliografice) de pe internet fără a fi nevoie ca acestea să fie introduse manual de către utilizator. În baza acestor facilităţi, procesul clasic de a redacta bibliografia unui articol printr-un efort asiduu cu durata de câteva ore este înlocuit de un demers inteligent,

care permite redactarea integrală a bibliografiei conform standardelor dorite (APA, MLA etc.) în numai câteva minute (*sic*!).

8.1.9. Anexe/materiale suplimentare disponibile online

Uneori, anumite aspecte ce ţin de secţiunea metodologică sau de partea de rezultate pot fi considerate ca fiind nepotrivite pentru a fi incluse direct în secţiunile respective, fie din cauza lungimii lor, fie din cauza faptului că afectează coerenţa/fluenţa textului, fie din cauza ambelor aspecte menţionate. Totuşi, dacă acestea sunt relevante pentru cititori şi pentru evaluatorii critici ai articolului respectiv, ele pot fi incluse într-una sau mai multe anexe (ca parte integrantă din articol, inclusiv în forma sa tipărită) sau ca material suplimentar disponibil online (adiacent versiunii electronice a articolului, dar absent din forma tipărită a acestuia; APA, 2010). Fără a se limita la elementele menţionate mai jos, în anexe pot fi cuprinse: (a) stimulii experimentali utilizaţi sau detalii despre orice altă probă/sarcină utilizată, în contextul în care nu există referinţe bibliografice accesibile pentru acestea, cu respectarea drepturilor de autor; (b) o descriere detaliată a populaţiei investigate; (c) anumite instrumente/facilităţi dezvoltate (cum ar fi sintaxe din SPSS etc.). În cazul materialelor online suplimentare pot fi incluse chiar mai multe informaţii decât în anexe: de la analize statistice detaliate până la baza de date a studiului.

8.2. Alte aspecte importante asociate publicării unui articol

În cea de-a doua parte a acestui capitol vom încerca să integrăm succint alte informaţii importante pentru procesul de redactare şi de publicare a unui articol ştiinţific, pornind de la câteva dintre cele mai frecvente întrebări ridicate de persoanele interesate să-şi publice contribuţiile editoriale în reviste de specialitate. În mod concret, vom încerca să răspundem la următoarele întrebări esenţiale: (a) „Ce persoane ar trebui incluse ca autori ai unui articol?”; (b) „Ce aspecte etice, asociate redactării şi prezentării articolului, trebuie respectate?”; (c) „Unde ar trebui să public articolul scris (în ce revistă)?”; (d) „Ce presupune procesul de evaluare colegială şi cum ar trebui să mă raportez la sugestiile/criticile primite?”.

8.2.1. Despre autoratul ştiinţific

Cel mai adesea, cercetările nu reprezintă demersuri singulare, ci eforturi colective în cadrul cărora, de multe ori, este greu să se diferenţieze, la nivelul cantităţii şi calităţii contribuţiilor, între cei care merită să fie autori ai studiului şi cei pentru care e suficientă menţionarea lor în nota autorilor, la mulţumiri.

Din acest punct de vedere, Browner (2006) specifică trei aspecte majore, dintre care, cel puţin unul[27] trebuie îndeplinit de un cercetător pentru a fi acceptat ca autor al studiului respectiv: (a) să aibă contribuţii intelectuale la dezvoltarea şi derularea studiului; (b) să participe la scrierea şi/sau revizuirea articolului; (c) să agreeze forma finală a articolului şi să-şi asume responsabilitatea publică pentru conformitatea

acestuia. Standardele APA (2010) menţionează că poate fi autor cel care are o contribuţie esenţială în derularea sau scrierea studiului şi care îşi asumă responsabilitatea pentru cele menţionate în studiu. Partea dificilă se referă la operaţionalizarea acestor aspecte menţionate.

Astfel, este greu de definit foarte clar ce înseamnă contribuţie intelectuală. Potrivit lui Browner (2006), aceasta s-ar putea regăsi: (a) fie în stabilirea întrebării de cercetat în urma studiului literaturii de specialitate şi a designului potrivit pentru a răspunde la aceasta (activitate de concepţie a studiului); (b) fie în prelucrarea statistică a datelor (verificarea ipotezelor studiului); (c) fie în intervenţii evaluative critice (activitate de analiză critică a rezultatelor studiului). După cum se poate deduce din această abordare, activităţile curente, de tip executiv, în care contribuţia se rezumă la a face doar ceea ce a fost trasat ca sarcină de lucru (de exemplu, recrutarea participanţilor, verificarea articolului din punct de vedere gramatical sau lingvistic, activitatea de testare/colectare a datelor şi de realizare a bazei de date; redactarea bibliografiei etc.) nu sunt considerate contribuţii intelectuale; prin urmare nu pot fi recompensate prin acordarea titulaturii de autor. Acest lucru este important de lămurit încă de la început, deoarece este posibil ca la realizarea cercetării să participe atât asistenţi de cercetare (implicaţi îndeosebi în colectarea datelor şi/sau construirea bazei de date), cât şi colegi şi sau mentori cu experienţă care se rezumă la simplul suport logistic, financiar sau la supervizarea activităţii, dar fără a aduce contribuţii ideatice noi). Acestora trebuie să li se comunice de la început statusul lor, pentru a evita conflictele. Desigur, între aceste prescripţii normative şi realitate pot exista discrepanţe izvorâte atât din relaţiile de putere (un asistent căruia i se sugerează să menţioneze ca autor un profesor/şef de departament/decan/conducător de doctorat, deşi persoana respectivă nu

are vreo contribuţie intelectuală), cât şi din înţelegeri reciproc avantajoase, dar care încalcă deontologia profesională („Eşti coautor în articolul meu, iar tu mă vei trece coautor în studiul tău, astfel că ne va creşte numărul publicaţiilor.”).

În ceea ce priveşte contribuţia unei persoane la scrierea articolului, ea poate fi adesea utilizată drept criteriu pentru stabilirea ordinii autorilor. Cutuma întâlnită în domeniul psihologiei, precum şi al altor ştiinţe conexe (ştiinţele educaţiei, medicină etc.) este aceea ca primul autor al unui articol să fie cel care scrie prima versiune a articolului şi care, de obicei, este cel care coordonează/integrează contribuţiile celorlalţi autori. Ordinea celorlalţi autori este proporţională cu contribuţia fiecăruia la articol (de exemplu, cel de-al doilea autor aduce o contribuţie mai importantă la articol decât al treilea etc.). O excepţie de la acest caz este situaţia în care toţi autorii provin din aceeaşi echipă, situaţie în care ultimul autor este de obicei supervizorul (conducătorul ştiinţific de doctorat).

În privinţa ultimei componente menţionate, cea a agreării formei finale a articolului şi a asumării responsabilităţii pentru cele prezentate în articol, ea a devenit din ce în ce mai importantă în ultimul timp, îndeosebi în urma scandalurilor de fraudă (de exemplu, falsificarea datelor) izbucnite periodic în ultimii ani (vezi detalii în Marshall, 1996, sau, mai recent, în Barlett, 2011, cu referire la cel mai recent caz din domeniul psihologiei sociale, cel al olandezului D. Stapel). O altă problemă frecvent întâlnită este aceea a prezentării trunchiate a rezultatelor (variabilelor), menţionându-se în articol doar cele care susţin punctul de vedere dorit – vezi Simmons, Nelson şi Simonshon (2011). Desigur, în astfel de cazuri nedorite se pune problema stabilirii responsabilităţii. Toţi autorii menţionaţi sunt de condamnat, chiar dacă unii dintre ei nu aveau

cunoştinţă de practicile respective? Pentru a rezolva această dilemă, unele reviste din domeniul medical au introdus pe lângă ideea de autorat, ideea de contributor (*contributorship*), respectiv de garant (*guarantors*) (Horton, 2008). Astfel, la finalul studiului, pe lângă nominalizarea autorilor se fac menţiuni cu privire la contribuţia fiecăruia dintre ei la studiu şi cu privire la persoanele responsabile pentru cele prezentate. Este posibil ca acest model aflat la început să fie preluat şi implementat pe scară largă şi de către alte reviste de specialitate, inclusiv din domeniul psihologiei.

Caseta 8.5. *Ilustrare ipotetică a contribuţiei şi garanţiei în studiul lui Sava* et al. *(2009)*

Autori: Sava, Yates, Lupu, Szentagotai şi David *Contribuţie*: Sava a scris manuscrisul şi a analizat statistic datele din perspectiva raporturilor cost-eficienţă şi cost-beneficiu. Yates s-a implicat în strategia de stabilire a costurilor implicate de cele trei intervenţii şi a contribuit la revizuirea articolului. Lupu şi Szentagotai au contribuit la designul studiului şi au coordonat procesul de colectare a datelor de pe teren. David a contribuit la designul studiului, la analiza efectelor celor trei tipuri de intervenţii şi la revizuirea articolului. *Garanţi*: Sava şi David.

8.2.2. Aspecte etice asociate procesului de redactare a articolului

Despre domeniul eticii în cercetare ar putea fi scrise mai multe volume de specialitate. Ele pot trata teme variate care merg de la conceperea studiului (existenţa unei aprobări din partea unui comitet de etica cercetării înainte de derularea unui studiu sau existenţa unui formular semnat de consimţire informată a participării la studiu dată de către subiecţii incluşi în cercetare) până la aspecte ce ţin de limita legalităţii (conflicte de interese) sau care o depăşesc (plagiatul, fabricarea sau falsificarea datelor).

În cele ce urmează mă voi referi doar la o parte dintre acestea, cele implicate în partea de raportare/publicare a rezultatelor. Dacă ne vom raporta la clasificarea lui Farthing (2008) cu privire la comportamentele inadecvate legate de raportarea rezultatelor unei cercetări, acestea ar putea fi împărţite în: (a) delicte minore, adică măsuri care „ajută” la obţinerea unor rezultate semnificative statistic (de exemplu, prezentarea trunchiată a rezultatelor, doar a celor care convin; eliminarea unor participanţi din studiu în baza unor decizii arbitrare şi, în principiu, orice „joc” cu datele sau decizie *post hoc* care să faciliteze obţinerea rezultatelor dorite, dar fără a apela la modificarea/fabricarea lor[28]); (b) şi fraude (plagiat, fabricarea rezultatelor, respectiv falsificarea rezultatelor). Le vom detalia pe cele din urmă.

Plagiatul constă în utilizarea unei părţi din munca publicată sau nepublicată a unei alte persoane şi în prezentarea acesteia ca fiind o contribuţie proprie, fără a menţiona prin referinţe bibliografice meritele autorului respectiv. Un caz aparte de fraudă îl reprezintă autoplagiatul, în cazul în care un autor prezintă ca informaţie nouă date expuse în studiile sale anterioare. Însă nu este suficientă doar menţionarea sursele bibliografice de inspiraţie: trebuie ca acest lucru să fie făcut cu respectarea drepturilor de autor. De exemplu, copierea unor pasaje mari de text (citarea unui text cu lungimea de o pagină cu precizarea referinţei bibliografice) nu reprezintă o conduită acceptabilă, chiar dacă a fost citată sursa din care a fost preluat materialul respectiv. Conform standardelor APA, pot fi preluate şi integrate direct în text citate de până la 40 de cuvinte. Pentru texte mai lungi, ele trebuie evidenţiate separat, ca paragrafe distincte (APA, 2010), iar dacă lungimea textului citat depăşeşte 400 de cuvinte ar trebui cerută permisiunea pentru reproducerea textului de la deţinătorii de drepturi ai acestuia. În toate aceste cazuri se adaugă ca

element al citării şi pagina din sursa de unde a fost preluat textul respectiv. O procedură similară ar trebui urmată în cazul figurilor sau a tabelelor. Politica APA permite, cu unele excepţii, preluarea a cel mult trei figuri sau tabele dintr-un articol sau capitol de carte, fără a cere permisiunea scrisă de la deţinătorii drepturilor, însă cel mai indicat ar fi să se verifice pentru fiecare caz în parte politica editorială cu privire la drepturile de autor (APA, 2010).

Pentru a evita asemenea situaţii neplăcute, a căror apariţie poate fi neintenţionată, se recomandă: (a) reformularea ideilor sau afirmaţiilor preluate din sursele citate, într-o proporţie de cel puţin 50%[29]; (b) citarea sursei bibliografice respective, pentru a nu fi acuzaţi de plagiat (vezi caseta 8.6).

Caseta 8.6. *Ilustrarea unei reformulări/parafrazări adecvate*

Pasaj original din Sava (2011):
Din punctul nostru de vedere, este prematur să ne pronunţăm asupra potenţialului utilizării măsurilor implicite ca detector de minciuni.

Reformulare insuficientă:
În opinia noastră, este prematur să ne pronunţăm asupra potenţialului de aplicare a măsurilor implicite ca detector de minciuni.

Reformulare corectă:
Sava (2011) consideră drept insuficiente dovezile ştiinţifice pentru a decide dacă tehnicile implicite pot fi utilizate pentru a detecta opiniile disimulate ale indivizilor.

În privinţa autoplagiatului, cea mai gravă situaţie este aceea de publicare redundantă, adică aceea de a publica două sau mai multe articole/cărţi etc. care să conţină integral sau într-un procent foarte mare aceleaşi informaţii şi fără a face referire una la alta, astfel încât ultima publicaţie pare a fi una originală. În strânsă legătură cu această practică se

regăseşte conceptul de utilizare corectă (*fair use*). În acest context, autorii pot reproduce, din lucrări personale anterioare, un fragment de text ce poate merge până la câteva paragrafe, dacă sunt respectate drepturile de autor (care, cel mai frecvent, aparţin editurii în care s-a publicat iniţial), dacă este menţionată explicit sursa primară, iar textul reprodus este încadrat ca citat (între ghilimele).

Apropiat ideii de autoplagiat, cât şi ideii de plagiat, în general, este conceptul de lucrare derivată (*derivative work*). O lucrare este considerată derivată dacă constă într-o reinterpretare, reelaborare etc. a unei lucrări deja existente, astfel încât scopul şi conţinutul noului text să fie suficient de diferit de sursa primară[30] (APA, 2010).

Un alt concept relaţionat cu această practică este principiul circulării limitate a informaţiei (*limited circulation principle*), prin care se permite publicarea unui articol apărut iniţial într-un mediu limitat ca accesibilitate (un document tipărit în puţine exemplare, într-o revistă cu caracter local etc.), într-un mediu cu accesibilitate largă (de exemplu, o revistă prestigioasă, de largă accesibilitate), dacă sunt respectate drepturile de autor şi dacă este menţionată sursa iniţială a materialului republicat.

O altă situaţie care trebuie evitată, chiar dacă nu constituie plagiat, este aceea în care setul de variabile este manipulat astfel încât mai multe articole sunt scrise ca fiind aparent independente, deşi folosesc acelaşi lot de persoane şi, mai ales, nu fac referire la articolele anterioare publicate, bazate pe acelaşi eşantion. Astfel, în caseta 8.7, cazul A reprezintă un demers onorabil, acceptat ca practică de lucru, în timp ce cazul B este un caz de conduită lipsită de etică, ce trebuie evitată de către cercetători, situaţie cunoscută în literatura de specialitate ca *salami publication/slicing/science*. De aceea, pe lângă obligativitatea menţionării explicite a situaţiilor în care o publicaţie nouă a fost realizată folosind

aceeiaşi baze de date care a fost utilizată pentru redactarea unui articol anterior, este important să se aibă în vedere criterii obiective pentru a redacta articole multiple folosind aceeaşi bază de date, cum ar fi existenţa unor obiective diferite ale studiilor. Astfel, poate fi normal să se abordeze problema demonstrării existenţei unei relaţii cauzale (studii cauzal-descriptive) separat de aceea a discutării mecanismelor cauzale implicate. În schimb, nu se justifică redactarea a două articole în baza unui studiu randomizat controlat, ambele referindu-se la acelaşi lucru, demonstrarea eficienţei unei forme de intervenţie, doar pentru simplul fapt că au fost utilizate două măsurători ale variabilei dependente primare (câte un studiu pentru fiecare variabilă dependentă în parte).

Caseta 8.7. *Ilustrarea unor practici acceptate, respectiv lipsite de etică cu privire la publicaţiile multiple obţinute folosind aceeaşi bază de date*

Cazul A – Acceptabil

În urma derulării unui studiu clinic controlat sunt redactate trei studii: unul care vizează eficienţa intervenţiei (asupra depresiei ca variabilă de interes primar, dar şi asupra calităţii vieţii ca aspect secundar urmărit), un altul care vizează mecanismele schimbării şi un al treilea studiu care urmăreşte raportul cost-eficienţă. Mai mult, în studiile publicate ulterior se face referire explicită la faptul că acestea se bazează pe date prezentate într-o publicaţie anterioară, acest aspect fiind subliniat în nota autorului.

Cazul B – Incorect

În urma derulării aceluiaşi studiu controlat sunt redactate cinci studii care vizează: (a) eficienţa intervenţiei focalizate pe variabila dependentă primară – depresia; (b) eficienţa intervenţiei asupra unei variabile dependente secundare, calitatea vieţii, dar prezentată în acest studiu secundar ca variabilă dependentă primară; (c) un studiu privind raportul cost-eficienţă, luând în calcul depresia; (d) un studiu privind raportul cost-utilitate luând în calcul calitatea vieţii; (e) un studiu ce vizează mecanismele schimbării implicate. Mai mult, pe lângă fragmentarea excesivă a studiilor, nici nu se face referire în vreunul dintre ele la publicaţii anterioare care utilizează aceeaşi bază de subiecţi, dând impresia cititorilor că ar fi vorba despre cinci studii independente.

Fabricarea datelor se referă la redactarea unui articol în baza unor date inexistente, „inventate”. Acest tip de fraudă este mai puţin întâlnit în comunitatea ştiinţifică de nivel înalt (totuşi, vezi cazul, descoperit, recent de fabricare a datelor de către un psiholog, Marc Hauser, descris în Wade, 2010), însă estimez că practica este destul de răspândită printre studenţi, acestora fiindu-le mai comod să inventeze rezultate statistice decât să urmeze întregul proces laborios de colectare a datelor în laborator/pe teren. Detectarea acestei practici este dependentă de versatilitatea autorilor ei. Astfel, asemenea tentative pot fi detectate destul de uşor, prin consultarea parametrilor statistici obţinuţi (de exemplu, valori exagerate ale testului t student, medii sau abateri standard nejustificat de diferite de cele obţinute în alte studii etc.), dacă autorii nu au nivel de expertiză statistică ridicat. În schimb, dacă există expertiză statistică şi metodologică ridicată, cazurile devin mult mai greu de detectat, ele putând fi evidenţiate mai degrabă prin absenţa probelor (dovezilor că datele au fost colectate).

Falsificarea datelor are în vedere modificarea deliberată a acestora (de exemplu, creşterea cu două puncte a scorurilor celor din grupul experimental sau „atât cât este nevoie” ca diferenţele dintre grupele comparate să devină semnificative statistic) sau a modalităţii de cotare a acestora (de exemplu, dacă iniţial grupul experimental era notat cu 1, iar cel de control cu 0, iar rezultatele obţinute contravin direcţiei menţionate în ipoteză, se inversează deliberat semnificaţia acordată codurilor, sub paravanul unor erori de codare, în cazul în care acestea sunt identificate). Asemenea modificări sunt foarte greu de identificat (deoarece există dovezi privind colectarea datelor), însă şi în cazul unor asemenea practici detectarea este dependentă de expertiza statistică a celor care comit asemenea infracţiuni.

Asemenea practici sunt destul de răspândite în toate domeniile ştiinţifice. În acest sens, o metaanaliză realizată de Fanelli (2009) indică faptul că 2% dintre cercetători au recunoscut sub protecţia anonimatului că au comis fapte grave, precum fabricarea sau falsificarea rezultatelor, iar o treime au recunoscut că au comis cel puţin o dată delicte minore. Procentul acestor conduite inadecvate a crescut dacă întrebarea era formulată în termeni impersonali („Cunoaşteţi pe cineva care…?”): 14% pentru situaţiile de tipul fabricării, falsificării sau modificării datelor, respectiv la 72% (*sic*!) în privinţa celorlalte practici considerate delicte minore.

De aceea, pentru a preveni şi combate asemenea practici, precum şi pentru a identifica vinovaţii pentru acest lucru[31], ar fi indicat să se ţină cont de câteva reguli simple: (a) asumarea statutului de autor doar dacă acesta este meritat; (b) menţionarea explicită la finalul articolelor a contribuţiei fiecărui autor şi a garantului (garanţilor); (c) lucrul în echipă la partea de analiză a datelor; (d) efectuarea doar a acelor analize statistice stipulate în planul iniţial (evitarea „jocului cu datele pentru a vedea ce iese”); (e) respectarea contribuţiilor anterioare prin asumarea (citarea) surselor relevante şi reformularea cu cuvinte proprii a celor citate; (f) asumarea articolului, în varianta finală, de către toţi autorii.

8.2.3. Alegerea revistei unde se va publica articolul

Decizia cu privire la locul unde va fi publicat articolul finalizat este una extrem de importantă. Ideal ar fi ca prima opţiune privind publicarea articolului să fie în mintea celui care redactează articolul încă de la început pentru a se plia pe specificul revistei respective. În realitate,

decizia se ia de multe ori după finalizarea redactării articolului, la care se adaugă cazurile în care articolul este respins de către revista avută iniţial ca primă opţiune.

Până la urmă, decizia de a alege o anumită publicaţie se bazează pe criteriul considerat cel mai important de către autori: rapiditatea publicării *vs* prestigiul publicării.

Rapiditatea publicării este facilitată de: (a) alegerea unor reviste cu acces liber (*open access*)[32]; (b) alegerea unor reviste care au un procent mai ridicat de acceptare a manuscrisului trimis spre publicare (de obicei însă, acest lucru corelează invers proporţional cu prestigiul revistei, cu excepţia revistelor cu acces liber); (c) alegerea unei reviste cu un volum mare de articole publicate într-un an de zile. În schimb, prestigiul publicării este facilitat de selectarea unor reviste cu un factor de impact ridicat în domeniu/scor de influenţă relativ mare.

Caseta 8.8. *Principalii indicatori scientometrici în aprecierea calităţii unei reviste*

Factorul de impact al revistei (Amin şi Mabe, 2007)

Acesta reflectă media citărilor în anul de referinţă (de exemplu, în anul 2012) ale articolelor apărute în ultimii doi ani anteriori celui de referinţă (articolele apărute în revista respectivă în anii 2010 şi 2011). Factorul de impact este calculat de Thomson Reuters şi este disponibil sub denumirea de *impact factor* în secţiunea Journal Citation Reports a celei mai prestigioase baze de date internaţionale Thompson Reuters (care dispune şi de ISI Web of Science). Începând cu anul 2007, pe lângă factorul de impact pentru intervalul de doi ani se poate calcula şi un factor de impact pentru un interval de cinci ani. În ambele cazuri, există două limite importante ale acestor indicatori: (a) valoarea medie a factorilor de impact variază de la un domeniu la altul (de exemplu, sunt mai mari în domeniul neuroştiinţelor decât în matematică, iar acest fapt nu permite compararea revistelor din domenii diferite pe baza factorului de impact); (b) se are în vedere doar numărul de citări primite, fără a include ca factor de ponderare calitatea revistelor în care apar citările (deşi pot exista mari diferenţe calitative între reviste). Revistelor indexate recent în ISI Web of Knowledge (Science), cărora nu li se poate calcula factorul de impact, li se atribuie un factor de impact egal cu zero.

Scorul eigenfactor (Whitton, 2011)
Acest indicator constituie un factor de impact al revistei ajustat pe cinci ani de zile, care încearcă să elimine cel de-al doilea neajuns menţionat, ponderând diferit citările în funcţie de calitatea revistelor din care provin. Astfel, ajustarea numărului de citări primite se face: (a) acordând o importanţă mai mare citărilor în reviste cu factor de impact ridicat şi celor în reviste a căror pagină de internet este accesată mai mult timp (considerat a fi un indicator al importanţei revistei); (b) eliminând autocitările (citările unor articole din revistă în alte articole din aceeaşi revistă fiind eliminate); (c) extinzând sfera de urmărire a citărilor de la domeniul ştiinţelor sociale (pentru psihologie) la toate domeniile ştiinţifice. Aşadar, *scorul eigenfactor*, spre deosebire de factorul de impact, ajustează calitativ numărul mediu de citări, însă păstrează aceeaşi limită a faptului că există variaţii între domeniile de studiu, fapt ce nu permite compararea revistelor din domenii diferite. În schimb, valoarea acestuia reprezintă influenţa revistei în general, nu impactul mediu al unui articol din revista respectivă. Pentru a corecta acest neajuns (şi pentru a obţine un indicator comparabil cu factorul de impact ce se referă la influenţa unui reviste per articol), se calculează scorul relativ de influenţă (*Article Influence Score*) care are avantajul de a fi o măsură standardizată în care un scor de 1 indică un impact mediu, un scor subunitar o influenţă mai scăzută, iar un scor supraunitar o influenţă sporită.

Scorul relativ de influenţă (Whitton, 2011)
Acest scor reprezintă raportul dintre factorul de impact al unei reviste şi media factorilor de impact ai revistelor din acelaşi domeniu. Dacă o revistă este inclusă în mai multe domenii de studiu, se ia în considerare domeniul de referinţă cu cea mai mică medie a factorilor de impact. Avantajul acestui indicator este că permite comparaţia între domenii (de exemplu, o revistă cu factor de impact 1,5, în contextul în care media factorilor de impact pentru revistele din domeniul respectiv este 1,0, va avea un scor relativ de influenţă de 1,5; acelaşi scor relativ de influenţă de 1,5 ar putea fi obţinut şi de o revistă dintr-un alt domeniu, dacă factorul de impact al acesteia este 3,0, iar factorul de impact agregat [media domeniului respectiv] este 2,0). Un alt avantaj major al acestui indice comparativ cu factorul de impact clasic este calitatea citărilor obţinute. Conform versiunii clasice, două reviste pot avea un factor de impact identic, dacă obţin acelaşi număr de citări/număr de articole publicate, chiar dacă prima revistă este citată numai în alte reviste de top (ce strâng foarte multe citări), iar cea de-a doua este citată numai în reviste din partea inferioară a clasamentului (cu puţine citări ulterioare). Scorul relativ de influenţă corectează acest neajuns şi are în plus faţă de *scorul eigenfactor* posibilitatea de a permite comparaţii între domenii, scorul 1 indicând un impact mediu asupra domeniului. O variantă derivativă a acestui indice este folosită ca indicator scientometric de bază în granturile acordate de CNCS/ANCS.

Recomandări generale privind alegerea revistei:

- este preferabil să vă adresaţi revistelor care abordează mai frecvent tematica discutată de dumneavoastră în articol; acest lucru se poate face simplu: să presupunem că studiaţi problematica angajamentului organizaţional şi luaţi în considerare trei reviste posibile A, B şi C; dacă tematica respectivă a fost abordată în ultimii ani mai frecvent în revista A decât în revistele B sau C, atunci ar fi de preferat să alegeţi revista A, care a dovedit preocupare în politica sa editorială pentru tema respectivă; în mod similar, puteţi alege revista în funcţie de similaritatea dintre cele mai importante referinţe bibliografice din studiul dumneavoastră şi revistele care au citat frecvent aceleaşi referinţe;
- este preferabil să selectaţi revistele în urma unei atente analize a colegiului redacţional şi a instrucţiunilor pe care le oferă autorilor; de exemplu, dacă autorii citaţi de către dumneavoastră fac parte din colegiul redacţional al revistei respective şi/sau publică frecvent în această revistă, respectiv dacă teoriile susţinute în articolul respectiv constituie o temă de ultimă oră (*hot issue*) pentru revista respectivă, atunci aceasta reprezintă o opţiune importantă; în mod similar, dacă manuscrisul se încadrează în cerinţele specificate în instrucţiunile către autori ale unei reviste din domeniul de interes (de exemplu, lungimea manuscrisului, cerinţele metodologice etc.), revista merită a fi luată în considerare pentru publicare;
- este preferabil să selectaţi revistele în funcţie de calitatea articolului realizat (atât din perspectivă metodologică, fiind mai apreciate studiile experimentale decât cele non-experimentale, cât şi din perspectiva impactului teoretic sau practic al articolului respectiv); dacă articolul

consonează cu politica editorială a unei reviste (de exemplu, este vorba de o metaanaliză, iar politica editorială a revistei menţionează explicit preferinţa pentru evaluări sistematice ale literaturii de specialitate), atunci revista respectivă ar putea constitui o opţiune serioasă; de asemenea, dacă articolul surprinde mai degrabă aspecte aplicative, dar riguroase metodologic, decât aspecte specifice cercetării fundamentale, ar fi indicat să căutaţi reviste care valorizează rigurozitatea metodologică şi au o orientare aplicativă, decât să vă orientaţi spre reviste de top care vizează cercetarea fundamentală;

- este preferabil să selectaţi revistele şi în funcţie de calitatea rezultatelor obţinute; de exemplu, un studiu în care rezultatele susţin foarte clar o anumită direcţie are mai mari şanse să fie publicat în reviste cu factor de impact ridicat decât un studiu în care rezultatele obţinute sunt inconsistente sau mai puţin robuste; în aceste din urmă cazuri este preferabil să fie selectate reviste cu un factor de impact mai scăzut;
- este preferabil să selectaţi reviste cu o rată de respingere a articolelor (în urma procesului de evaluare colegială) mai scăzută, dacă sunteţi interesat să publicaţi cât mai repede studiile sau dacă acestea reprezintă continuări ale unor studii deja publicate pornind de la aceeaşi bază de date;
- este preferabil să selectaţi reviste de top, cu o calitate ridicată a articolelor, dacă aveţi încredere că articolul va aduce o contribuţie nouă şi însemnată în domeniul respectiv şi dacă doriţi să maximizaţi şansele ca articolul să nu treacă neobservat ulterior (să crească impactul studiului în domeniul de referinţă).

Desigur, preferabilă ar fi publicarea în reviste cotate ISI (cu factor de impact), fie ele dintre cele bazate pe o politică editorială clasică sau reviste cu acces deschis. Dacă acest lucru nu este posibil, există şi alte variante de publicare, cum ar fi aceea de a publica în reviste indexate în baze de date internaţionale (cum ar fi PsycINFO, Scopus, ERIC etc.), deşi nu sunt indexate ISI Web of Science. Asemenea reviste sunt potrivite mai ales pentru cei care se află la început de carieră sau în cazul unor articole corect formulate, dar cu un impact redus la dezvoltarea cunoaşterii din domeniu.

Pe scurt, dacă se porneşte de la premisa că nu este important doar să se publice un articol (să existe o publicaţie de specialitate), ci contează în primul rând ca studiul publicat să aibă un impact în dezvoltarea domeniului de interes, este esenţial să se publice articole de specialitate doar în reviste indexate în baze de date internaţionale, dintre acestea preferabile fiind articolele indexate ISI Web of Knowledge (Science) cu factor mare de impact, deoarece oferă cele mai mari şanse ca ideile articolului să fie preluate şi citate ulterior. Dacă aceasta este ţinta finală, cu referire la publicaţiile periodice din domeniul psihologiei, trebuie analizat acest deziderat prin prisma aspectelor menţionate anterior, deoarece revistele cu un factor de impact ridicat în domeniul lor de studiu sunt şi cele în care este cel mai greu să ajungi să publici, existând o rată mare de respingere a manuscriselor trimise spre publicare.

Tabelul 8.7. *Exemplificarea indicatorilor scientometrici pentru reviste din sfera psihologiei*

Domeniul/denumirea revistei	Tipul revistei	Factor de impact (doi ani)	Factor de impact (cinci ani)	Eigenfactor	Scor relativ de influenţă*	Locul de clasare în domeniu
Domeniul psihologiei clinice						

Journal of Consulting and Clinical Psychology	Top	5,22	6,20	0,0238	3,804	4/104
Journal of Clinical Psychology	Mediu	1,61	2,14	0,0073	1,006	47/104
American Journal of Family Therapy	Scăzut	0,41	0,51	0,0004	0,340	91/104
Domeniul psihologiei aplicate						
Journal of Applied Psychology	Top	3,97	6,73	0,0313	4,337	3/69
Ergonomics	Mediu	1,37	1,78	0,0054	1,147	30/69
Journal of Employment Counseling	Scăzut	0,41	0,41	0,0002	0,143	60/69
Domeniul psihologiei educaţionale						
Journal of Educational Psychology	Top	3,58	4,92	0,0123	3,266	2/50
British Journal of Educational Psychology	Medie	1,01	2,32	0,0033	1,391	27/50
European Journal of Psychology of Education	Scăzut	0,54	0,95	0,0008	0,420	44/50
Domeniul psihologiei sociale						
Journal of Personality and Social Psychology	Top	5,20	7,43	0,0541	5,577	3/58
Journal of Social and Clinical Psychology	Mediu	1,29	2,01	0,0043	1,328	26/58
Journal of Social Psychology	Scăzut	0,69	1,10	0,0018	0,689	50/58
Domeniul psihologiei experimentale						

Journal of Cognitive Neuroscience	Top	5,35	6,28	0,0395	3,000	3/81
Memory & Cognition	Mediu	1,79	2,45	0,0146	1,127	42/81
Perceptual and Motor Skills	Scăzut	0,49	0,57	0,0032	0,157	77/81

* Toţi indicatorii sunt valabili pentru anul 2010, excepţie făcând datele ce privesc scorul relativ de influenţă pentru care valorile se referă la anul 2009 (sursa: Thomson Reuters [n.d.]). Revistele din sfera psihologiei sunt grupate în domenii specifice (de exemplu, psihologie clinică), în tabelul 8.7 fiind trecute doar o parte dintre aceste domenii.

8.2.4. Recomandări cu privire la procesul de evaluare colegială

Aşa cum am precizat în primul capitol al cărţii, una dintre caracteristicile esenţiale ale unui demers ştiinţific este aceea că orice studiu este supus unui proces de evaluare colegială (*peer review*), al cărui scop este de: (a) a decide dacă un articol poate fi publicat, în funcţie de exigenţele specifice ale fiecărei reviste (fiind preferate articole corecte metodologic, scrise clar şi cu impact [contribuţie] semnificativ[ă] sub aspect teoretic sau practic); (b) a oferi, în cazul unui răspuns afirmativ la punctul (a), un feedback critic constructiv care să conducă la îmbunătăţirea formei finale a articolului trimis spre publicare.

Manualul APA (2010) sintetizează principalele tipuri de răspunsuri pe care le poate primi un autor pentru articolul trimis spre publicare:

(a) *acceptarea* (acceptare cu modificări minore) reprezintă cel mai fericit caz, însă întâlnit rar, de a publica articolul trimis, eventual după unele modificări punctuale; o asemenea decizie este întâlnită ceva mai des în cazul revistelor cu factor de impact scăzut sau indexate în alte baze de date internaţionale decât ISI Web of Science;

(b) *respingerea* reprezintă cel mai nefericit caz: articolul este respins fără posibilitatea de a-l retrimite într-o variantă revizuită (acest caz apare atunci când articolul nu se află în sfera de preocupare a revistei, conţine deficienţe teoretice sau metodologice majore care nu pot fi corectate prin revizuirea articolului, este scris într-o formă inacceptabilă sau, cel mai adesea, când nu aduce vreo contribuţie la dezvoltarea cunoaşterii sau are o contribuţie considerată nesemnificativă; această decizie este cel mai frecvent întâlnită în cazul revistelor cu factor de impact ridicat;

(c) *respingerea cu posibilitatea de a retrimite o variantă revizuită* (*revise and resubmit*) – este cea mai întâlnită decizie pentru revistele cotate ISI cu factor de impact mediu; acest lucru înseamnă că lucrarea are potenţialul de a contribui la dezvoltarea domeniului studiat, dar că manuscrisul trimis conţine încă multe aspecte ce trebuie explicitate pentru a lua o decizie finală în acest sens şi pentru a valorifica la maximum potenţialul articolului respectiv.

Dincolo de aceste considerente generale, vom formula câteva aprecieri şi recomandări utile pentru succesul procesului de evaluare colegială, succes operaţionalizat prin publicarea articolului trimis în acest scop.

În primul rând, trebuie să se aibă în vedere faptul că nu este vorba despre un proces de evaluare a calităţii autorilor, ci despre evaluarea calităţii articolului. Prin urmare, este posibil ca autori prestigioşi să aibă articole respinse sau să primească o decizie care să solicite o revizuire majoră a articolului înainte de a-l trimite din nou spre evaluare revistei respective. De exemplu, o statistică privind situaţia din anul 2010 (APA, 2011) indică faptul că rata de respingere a manuscriselor trimise pentru

publicare la reviste de top din domeniul psihologiei este în jurul a 80% (88% pentru *Journal of Applied Psychology*; 81% pentru *Journal of Personality and Social Psychology*; 79% pentru *Journal of Educational Psychology* etc.). De aceea, în astfel de cazuri, chiar articole de calitate pot fi respinse dacă nu se precizează suficient de clar în conţinutul lor impactul pe care îl pot avea la nivel teoretic sau practic.

În consecinţă, fiind vorba despre exigenţe sporite privind calitatea manuscrisului trimis spre publicare, concentrarea autorilor ar trebui să fie în direcţia oferirii de răspunsuri punctuale la întrebările evaluatorilor, chiar dacă feedbackurile primite de la evaluatori nu sunt unele foarte încurajatoare. Perioada standard în care este foarte probabil să primim un răspuns pentru un manuscris trimis este cuprinsă adesea între două până la trei luni de zile. Aceeaşi perioadă de timp este oferită autorilor care au primit recomandarea de revizuire şi retrimitere a textului, respectiv de acceptare cu modificări minore, pentru a răspunde la solicitările primite. Este important să se aloce suficient timp pentru a pregăti cele mai bune răspunsuri, iar acestea trebuie oferite punctual, pentru fiecare problemă ridicată în urma evaluării colegiale.

În privinţa răspunsurilor, Budgell (2009) sugerează autorilor să adopte un comportament de acceptare a recomandărilor şi sugestiilor primite pentru a creşte probabilitatea de a primi acceptul de publicare a studiului respectiv. Acest lucru este mai cu seamă valabil pentru recomandările venite din partea editorului[33] sau pentru cele provenite din partea a cel puţin doi dintre evaluatorii desemnaţi. Atunci când nu suntem de acord cu sugestiile oferite, răspunsurile trebuie să fie formulate punctual, în termeni politicoşi. Omiterea unor sugestii, precum şi răspunsurile de tip ironic nu sunt apreciate de către evaluatori.

În demersul de evaluare colegială pot fi parcurse mai multe asemenea cicluri până la luarea unei decizii finale de acceptare sau respingere a unui articol. Nu întotdeauna vom avea desemnaţi aceiaşi evaluatori pentru articol; de aceea, este posibil să primim noi sugestii de modificare a articolului, inclusiv în etape ulterioare evaluării.

O altă recomandare vizează politica unor reviste de specialitate de a permite autorilor să nominalizeze persoanele care vor fi desemnate ca evaluatori. Dacă acest lucru este permis, Budgell (2009) consideră că ar trebui să se recurgă la această nominalizare, mai ales dacă aveţi încredere în obiectivitatea/utilitatea recomandărilor provenite de la anumiţi evaluatori, acest fapt sporind calitatea manuscrisului trimis spre publicare.

De asemenea, este preferabil ca partea de discuţii să fie scrisă într-un limbaj relativist, fără a emite judecăţi finale cu privire la susţinerea sau lipsa de susţinere pentru o anumită teorie. Din acest punct de vedere, o atitudine diplomată este mai apreciată decât formularea unor afirmaţii tranşante care să vină în contradicţie cu opiniile unor evaluatori.

Dacă manuscrisul a fost respins, înainte de a-l trimite către o altă revistă ar fi util să integraţi în text sugestiile primite pe care le consideraţi îndreptăţite, deoarece acest lucru va duce la îmbunătăţirea articolului dumneavoastră. Prin urmare, trataţi inclusiv manuscrisele respinse ca şi cum ar fi primit decizia de revizuire şi retrimitere.

O altă recomandare este aceea de a vă specializa (aprofunda) pe un domeniu de cercetare restrâns. Acest lucru vă va permite în timp, prin participarea la conferinţe, discuţii cu specialiştii etc., să cunoaşteţi mai îndeaproape nu numai publicaţiile majore apărute, ci şi modul de gândire şi interesele specifice ale principalilor actori din domeniu. Acest lucru vă va ajuta să vă construiţi mai bine partea de interpretare a rezultatelor şi de integrare a acestora în literatura de specialitate, deoarece veţi putea

răspunde mai punctual la preocupările şi perspectivele specialiştilor principali pe zona tematicii abordate.

În sfârşit, o ultimă recomandare vizează acceptarea schimbării rolurilor, atunci când vi se oferă ocazia, adoptând rolul de evaluator. Astfel este posibil să fiţi contactaţi inclusiv de editori necunoscuţi care să vă roage să evaluaţi un manuscris din zona dumneavoastră de interes. O asemenea experienţă vă va ajuta ulterior să vă puneţi mai uşor în postura colegului evaluator şi să înţelegeţi mai bine aşteptările acestuia.

MacAuley (2008) oferă o serie de sugestii pentru evaluatori; două dintre acestea le amintim în continuare. Astfel, un sfat util ar fi ca evaluatorul să citească mai mult despre tema abordată decât articolul supus evaluării, pentru a putea aprecia mai bine gradul de noutate adus de articol, contribuţia sa la stadiul cunoaşterii. Apoi, o altă recomandare vizează structurarea feedbackului în trei părţi distincte: (a) un comentariu general despre lucrare (cu referire la originalitate [noutate], importanţă şi la validitatea concluziilor formulate: „Aduce noutăţi acest studiu şi este el bine realizat?”); (b) discutarea problemelor majore identificate; (c) precizarea, preferabil în ordinea apariţiei lor în text, a problemelor minore sesizate.

O altă calitate pe care ar trebui să o posede atât evaluatorul, cât şi autorul evaluat este aceea de a dovedi o bună cunoaştere a metodologiei cercetării şi a modului în care diferite decizii legate de implementarea studiului afectează validitatea concluziilor formulate, mai cu seamă că din acest punct de vedere cele două părţi se află adesea pe poziţii contrare: autorul, încearcă să evidenţieze cât mai multe aspecte pozitive ale lucrării sale, pentru a creşte impactul acesteia, iar evaluatorul adoptă o poziţie sceptică şi sancţionează afirmaţiile lipsite de suportul empiric determinat de tipul de studiu şi de rezultatele obţinute. Din acest punct de vedere,

standardele menţionate în tabelul 8.2 merită luate drept criteriu de referinţă pentru a putea formula obiecţii sau răspunde la acestea, în cadrul procesului de evaluare colegială.

Concluzii

Scopul acestui capitol a fost de a oferi cititorilor interesaţi o serie de repere care să-i ajute să redacteze sau să evalueze articole de specialitate, în conformitate cu standardele internaţionale din domeniul psihologiei. În acest sens, am expus structura unui articol ştiinţific empiric, prezentând pe larg modelul IMRAD (introducere, metodologie, rezultate şi discuţii). Deşi este vorba despre o structură comună pentru diferite tipuri de articole, recomandarea noastră expresă este de a adapta conţinuturile fiecărei secţiuni apelând la standardele dezvoltate în acest sens. Astfel, pe lângă standardele JARS şi MARS dezvoltate de APA (2008), deosebit de utile sunt standardele consacrate în alte domenii de cunoaştere ştiinţifică, cum este cel al medicinei. Ne referim aici în primul rând la standardele dezvoltate de grupul CONSORT (pentru studii experimentale), însă merită analizate şi standardele pentru alte tipuri de studii, precum PRISMA (pentru studii metaanalitice bazate pe studii experimentale) sau STROBE (pentru diferite tipuri de studii non-experimentale). Pornind de la asemenea modele se pot redacta studii de specialitate care să răspundă anticipat la majoritatea exigenţelor cerute în procesul de evaluare colegială. Aceleaşi standarde pot fi utile şi în faza de concepere a studiilor.

Însă pentru a fi performanţi în domeniul cercetării ştiinţifice, trebuie dezvoltate şi alte abilităţi decât cele care vizează standardele de redactare

a unui articol empiric. M-am referit aici, în cea de-a doua parte a capitolului, la patru aspecte aparent distincte, dar care concurează la obţinerea succesului profesional în domeniul cercetării. Este vorba, în primul rând, de adoptarea unui spirit etic în cercetare, inclusiv în privinţa responsabilităţilor ce decurg din asumarea calităţii de autor. Apoi este vorba de abilitatea de a alege cea mai potrivită publicaţie pentru manuscrisul realizat şi de a interacţiona eficient cu editorul revistei şi cu evaluatorii articolului, pe parcursul întregului proces editorial.

Lecturi suplimentare

APA (2010). *Publication Manual of the American Psychological Association* (ediţia a VI-a). Washington DC: APA.

Într-un limbaj informal, am putea spune că reprezintă „biblia” cercetătorilor din domeniul psihologie. Lucrarea conţine referiri detaliate despre toate secţiunile acestui capitol, fiind cartea de referinţă care ar trebui consultată pentru diferite aspecte specifice ale cercetării (autorat, aspecte etice, standardul de redactare JARS etc.).

APA (2010). *Preparing manuscripts for publication in psychology journals: A guide for new authors.* Disponibil la http://www.apa.org/pubs/authors/new-author-guide.pdf

Conţine un material sintetic, redactat de un grup de cercetători cu experienţă sub girul APA, cu privire la cele mai importante aspecte în pregătirea manuscrisului şi în gestionarea acestuia în perioada de evaluare colegială.

Boutron I., Moher D., Altman D.G., Schulz K., Ravaud, P., for the CONSORT group (2008). Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*, 148, pp. 295-309.

Sistemul CONSORT este probabil cel mai avansat model de raportare a studiilor experimentale, fiind menţionaţi şi explicaţi paşii necesari de urmat pentru redactarea unor articole de calitate fundamentate pe studii randomizate. În articolul menţionat accentul este pus pe o anumită categorie de intervenţii (cele non-farmaceutice) aplicabile cazurilor specifice din psihologie sau psihoterapie.

Sava, F.A. (2005). „Cum să citeşti critic un articol?”. În Z. Bogathy, C. Sulea (ed.). *Manual de tehnici şi abilităţi academice* (pp. 37-45). Timişoara: Editura Universităţii de Vest.

Acest capitol scurt prezintă o grilă de lectură a articolelor empirice menită să întărească analiza critică a literaturii de specialitate şi capacitatea cititorilor de a identifica puncte tari sau slabe în cadrul unui articol redactat sau primit spre evaluare colegială.

Sternberg, R.J. (ed.) (2006). *Reviewing scientific works in psychology.* Washington DC: APA.

Cartea ar putea fi privită ca o extensie a modului în care este evaluată critic literatura de specialitate. Lucrarea conţine şapte capitole referitoare la adaptarea procesului de evaluare colegială pentru: studii empirice, articole metodologice, studii teoretice, evaluări sistematice ale literaturii de specialitate, cărţi, capitole de cărţi şi proiecte (granturi) de cercetare.

Bibliografie

American Educational Research Association (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35, pp. 33-40.

American Psychological Association (2010). *Publication Manual of the American Psychological Association* (ediţia a VI-a). Washington DC: APA.

American Psychological Association (2011). Summary report of journal operations, 2010. *American Psychologist*, 66, pp. 405-406.

Amin, M., Mabe, M. (2007). Impact factors: use and abuse. *Perspectives in publishing*, 1, pp. 1-6. Disponibil la adresa http://www.elsevier.com/framework_editors/pdfs/Perspectives1.pdf.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63, pp. 839-851.

Barlett, T. (2011). The fraud who fooled (almost) everyone. *The Chronicle of Higher Education*. Disponibil la adresa: http:// http://chronicle.com/blogs/percolator/the-fraud-who-fooled-almost-everyone/27917.

Boutron, I., Moher, D., Altman, D.G., Schulz, K., Ravaud, P., for the CONSORT group (2008a). Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*, 148, pp. w60-w66.

Boutron, I., Moher, D., Altman, D.G., Schulz, K., Ravaud, P., for the CONSORT group (2008b). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*, 148, pp. 295-309.

Browner, W.S. (2006). *Publishing and presenting clinical research* (ediţia a II-a). Philadelphia, PA: Lippincott Williams & Wilikins.

Budgell, B.S. (2009). *Writing a biomedical research paper. A guide to structure and style*. Tokio: Springer.

Campbell, M.K., Elbourne, D.R., Altman, D.G. (2004). CONSORT statement: extension to cluster randomised trials. *British Medical Journal*, 328, pp. 702-708.

Cooper, H. (2011). *Reporting research in psychology. How to meet journal article reporting standards*. Washington DC: APA.

Consiliul Naţional al Cercetării Ştiinţifice (n.d.). *Pachet de informaţii*. Disponibil la http://uefiscdi.gov.ro/userfiles/file/COMPETITIE%202011_IDEI/PCE/pachet_informatii%20PCE.pdf

Des Jarlais, D.C., Lyles, C., Crepaz, N., and the TREND Group (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, 94, pp. 361-366.

Equator Network (2012). *Enhancing the quality and transparency of health research*. Disponibil la http://www.equator-network.org/resource-centre/

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and metaanalysis of survey data. *PloS one*, 4, e5738.

Farthing, M.J.G. (2008). Ethics of publication. G.M. Hall (ed.), *How to write a paper* (ediţia a IV-a) (pp. 123-133). Oxford, UK: Blackwell Publishing.

Girden, E.R. (2001). *Evaluating Research Articles From Start to Finish* (ediţia a II-a). Thousand Oaks, CA: Sage Publications.

Hall, G.M. (2008). Structure of a scientific paper. În G.M. Hall (ed.), *How to write a paper* (ediţia a IV-a) (pp. 1-4). Oxford, UK: Blackwell Publishing.

Horton, R. (2008). Who should be an author? În G.M. Hall (ed.), *How to write a paper* (ediţia a IV-a) (pp. 112-116). Oxford, UK: Blackwell Publishing.

MacAuley, D. (2008). The role of the manuscript assessor. În G.M. Hall (ed.), *How to write a paper* (ediţia a IV-a) (pp. 92-102). Oxford, UK: Blackwell Publishing.

Marshall, E. (1996). Fraud strikes top genome lab. *Science*, 274, pp. 908-910.

Mendeley (n.d.). *Mendeley – free reference manager and pdf organizer*. Disponibil la http://www.mendeley.com

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., the PRISMA Group (2009). Preferred reporting items for systematic reviews and metaanalyses: The PRISMA Statement. *PLoS Med*, 6, e1000097.

Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., Altman, D.G., for the CONSORT Group (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. *British Medical Journal*, 340, c869.

Moss, F. (2008). Titles, abstracts, and authors. În G.M. Hall (ed.), *How to write a paper* (ediţia a IV-a) (pp. 34-41). Oxford, UK: Blackwell Publishing.

Nicol, A.A.M., Pexman, P.M. (2010). *Presenting your findings: a practical guide for creating tables* (ediţia a VI-a). Washington DC: APA.

Parmley, W.W. (1997). Authorship: taking the high road. *Journal of American College of Cardiology*, 29, p. 702.

Piaggio, G., Elbourne, D.R., Altman, D.G., Pocock, S.J., Evans, S.J.W. (2006). Reporting of noninferiority and nonequivalence randomized trials: An extension of the CONSORT statement. *JAMA*, 295, pp. 1152-1160.

Priebe, H.J. (2008). Results. În G.M. Hall (ed.), *How to write a paper*(ediţia a IV-a) (pp. 19-30). Oxford, UK: Blackwell Publishing.

Ramsey, S., Willke, R., Briggs, A., Brown, R., Buxton, M., Chawla, A., Cook, J., Glick, H., Liljas, B., Petitti, D., Reed, S. (2005). Good research practices for cost-effectiveness analysis alongside clinical trials: The ISPOR RCT-CEA task force report. *Value in Health*, 8, pp. 521-533.

Sava, F.A. (2005). Cum să citeşti critic un articol de specialitate? În Z. Bogathy şi C. Sulea (eds.), *Manual de tehnici şi abilităţi academice* (pp. 37-45). Timişoara: Editura Universităţii de Vest.

Sava, F.A., Yates, B.T., Lupu, V., Szentagotai, A., David, D. (2009). Cost-effectiveness and cost-utility of Cognitive Therapy, Rational Emotive Behavioral Therapy, and Fluoxetine (Prozac ®) in Treating Depression: A Randomized Clinical Trial. *Journal of Clinical Psychology*, 65, pp. 36-52.

Sava, F.A., Maricuţoiu, L.P., Rusu, S., Macsinga, I., Vîrgă, D. (2011). Implicit and explicit self-esteem and irrational beliefs. *Journal of Cognitive and Behavioral Psychotherapies*, 11, pp. 97-111.

Sava, F.A., Maricuţoiu, L.P., Rusu, S., Macsinga, I., Vîrgă, D., Cheng, C.M., Payne, B.K. (2012). An inkblot for the implicit assessment of personality: The semantic misattribution procedure. *European Journal of Personality*, 26(6), pp. 613-628.

Simmons, J.P., Nelson, L.D, Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, pp. 1359-1366.

Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine*, 152, pp. 1-7.

Sternberg, R.J. (2006). *Reviewing scientific works in psychology*. Washington, DC: APA.

Sternberg, R.J., Goordeeva, T. (1996). The anatomy of impact: What makes an article influential? În *Psychological Science*, 7, pp. 69-75.

Szentagotai, A., David, D. (2010). The efficacy of cognitive-behavioral therapy in bipolar disorder: a quantitative metaanalysis. *Journal of Clinical Psychiatry*, 71, pp. 66-72.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A, Thacker, S.B. (2000). Metaanalysis of observational studies in epidemiology: a proposal for reporting. Metaanalysis of observational studies in epidemiology (MOOSE) group. JAMA, 283, pp. 2008-2012.

Thompson, B. (1999). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Theory & Psychology*, 9, pp. 165-181.

Thomson Reuters (n.d.). Journal Citation Report. Disponibil la http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/

Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gøtzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J., Egger, M., STROBE initiative (2007). Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Annals of Internal Medicine*, 147, pp. w163-194.

Von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C, Vandenbroucke, J.P. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology*, 61, pp. 344-349.

Wade, N. (2010, 20 august). Harvard finds scientist guilty of misconduct. *The New York Times*, disponibil la adresa: http://www.nytimes.com/2010/08/21/education/21harvard.html.

Whitton, M. (2011). *Eigenfactor metrics – Eigenfactor score and article influence score*. Disponibil la adresa http://www.soton.ac.uk/library/research/bibliometrics/factsheet04-eigenfactors.pdf.

Zwarenstein, M., Treweek, S., Gagnier, J.J., Altman, D.G., Tunis, S., Haynes, B., Oxman, A.D., Moher, D. for the CONSORT and Pragmatic Trials in Healthcare (Practihc) group (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *British Medical Journal*, 337, a2390..

1. Structurarea unui articol de specialitate va ţine seama de publicul-ţintă căruia i se adresează. Astfel, dacă e vorba de un public format din specialişti în tema respectivă (de exemplu, pentru un articol care vizează impactul terapiei cognitiv-comportamentale în combaterea depresiei majore la adolescenţi, evaluat prin intermediul unui studiu clinic controlat, într-o revistă precum *Journal of Clinical and Consulting Psychology*), nu este nevoie să apelăm la vechiul sfat a lui Rousseau, potrivit căruia, înainte de toate, trebuie să ne definim termenii-cheie, cum ar fi *depresia majoră*, *studiul clinic controlat*, *terapia cognitiv-comportamentală*, *perioada adolescenţei*. Este suficient să apelăm la trimiteri bibliografice către studii care introduc/definesc conceptele respective. În schimb, într-un articol de popularizare adresat unui public mai puţin specializat, definirea termenilor-cheie este recomandabilă, deoarece acest demers i-ar ajuta pe cititori să înţeleagă mai bine despre ce este vorba în articol.

2. Această succesiune este cunoscută şi sub acronimul de IMRAD de la termenii în limba engleză *introduction, method, results* and *discussion*.

3. Se au în vedere titlurile formulate în limba engleză, recunoscută ca o limbă ce permite formulări concise. În limba română consider mai potrivită o limită de maximum 15 cuvinte.

4. . Din analiza titlurilor nominale se observă o apropiere foarte mare faţă de modelul PI(C)O, fiind regăsite în titlu informaţii despre populaţia investigată, intervenţia (sau aspectul central

urmărit) şi efectele acesteia (variabila dependentă principală, *primary outcome*).

5. Titlul în limba engleză este: „In moderation we trust: personality and teamwork skills” (o parafrazare a expresiei de pe bancnota de un dolar: In god we trust).

6. Titlul în limba engleză este: „The dark-side of teaching: a meta-analysis of the impact of teacher’s misbehavior on pupils”.

7. JARS este acronimul pentru sintagma din engleză *Journal Articole Reporting Standards*. Acestea au fost dezvoltate de un grup de lucru omonim, format din specialişti recunoscuţi (redactori-şefi ai unor reviste APA, indexate ISI Web of Science). Grupul a funcţionat sub egida APA (Asociaţia Psihologilor Americani), iar principalele rezultate ale grupului, inclusiv standardele JARS sunt prezentate în APA (2010) şi detaliate în Cooper (2011).

8. În rezumatele articolelor publicate în domeniul sănătăţii s-a răspândit un model structurat în şapte puncte, preluat din medicină: obiectiv, design, localizare, participanţi, intervenţii, variabile dependente principale şi concluzii.

9. Referinţele din studiul lui Sava et al. (2012) sunt menţionate doar în scop ilustrativ, ele nefiind incluse în bibliografia capitolului.

10. Limitele studiilor anterioare discutate sunt doar cele relevante pentru articolul de redactat, adică cele dependente de obiectivul studiului. De exemplu, dacă obiectivul studiul este stabilirea unei relaţii de tip cauză-efect, relevante sunt limitele studiilor anterioare care se referă la validitatea internă. Detalii despre tipurile de validitate şi implicaţiile acestora sunt prezentate în capitolul 3.

11. În acest sens, Asociaţia Americană a Cercetării Educaţionale (AERA, 2006) prezintă patru contribuţii-cheie în funcţie de obiectivul studiului: (a) pentru studii empirice care sprijină un model teoretic existent, se va sublinia contribuţia cercetării la elaborarea, consolidarea sau extinderea perspectivei teoretice respective; (b) dacă studiul sugerează un nou model explicativ (teoretic), trebuie să se arate ce este şi de ce este nevoie de această nouă perspectivă (vechile teorii/conceptualizări fiind inadecvate pentru a integra rezultatele studiului); (c) dacă studiul aduce contribuţii la nivel practic, aplicativ, ar trebui precizate aceste contribuţii şi relevanţa lor; (d) dacă obiectivul urmăreşte acoperirea unui deficit informaţional pe o anumită temă, ar trebui explicitat ce anume lipseşte şi de ce este important să se acopere acest gol de cunoaştere.

12. În tabelul 8.2 sunt prezentate doar o parte dintre standardele existente, selectate dintre cele considerate mai relevante din perspectiva paradigmei bazate pe dovezi empirice, validate ştiinţific. O listă mai cuprinzătoare de standarde pentru diferite tipuri de studii (peste 90 de standarde şi recomandări, inclusiv pentru studii calitative, studii descriptive de tipul sondajelor de opinie etc.), precum şi alte resurse utile sunt disponibile în Equator Network (2012) (http://www.equator-network.org/resource-centre/).

13. În cazul metaanalizelor, standardul de raportare este MARS (*Meta-Analysis Reporting Standards*), fiind elaborat de aceeaşi echipă care a elaborate şi JARS (Cooper, 2011).

14. Deşi ideea unui standard recomandat de redactare a unui articol în domeniul psihologiei trebuie salutată, existenţa unui singur cadru de referinţă pentru toate tipurile de studii constituie o

limită importantă a JARS în raport cu celelalte standarde. Chiar dacă JARS a fost dezvoltat modular (astfel încât să se ia în considerare particularităţile specifice pentru studiile experimentale faţă de cele cvasiexperimentale), mult mai potrivită ar fi fost strategia de a dezvolta standarde specifice pentru anumite categorii de studii (de exemplu, studii longitudinale, studii de caz etc. pentru care modulul general A1.1 ar putea fi mai puţin potrivit).

15. De cele mai multe ori, prin procedură se are în vedere procedura experimentală (a cercetării), adică partea de descriere a secvenţelor de colectare a datelor şi a contextului în care au fost colectate datele, astfel încât să fie posibilă replicarea studiului şi să existe un risc scăzut de distorsionare a datelor. Uneori însă, tot în zona de procedură sunt incluse unele prelucrări statistice preliminare pentru a obţine indicatorii necesari de analizat.

16. În privinţa designului, acesta fie este prezentat ca o componentă separată a metodologiei cercetării, situată înaintea prezentării participanţilor (situaţie tipică în cazul studiilor experimentale, a celor cvasiexperimentale, precum şi a studiilor non-experimentale longitudinale), fie este inclus implicit în subsecţiunea de procedură (situaţie specifică mai degrabă studiilor cu un design simplu, cum ar fi cazul studiilor corelaţionale – non-experimentale transversale). În cazul includerii designului în procedură, secţiunile majore rămân: participanţi, instrumente şi procedură (design). În caz contrar, succesiunea secvenţelor majore este: design, participanţi, instrumente şi procedură.

17. O trecere în revistă a tuturor riscurilor care pot afecta diferitele aspecte ale validităţii unui studiu este prezentată în capitolul 3.

18. O sursă bibliografică excelentă pentru redactarea tabelelor conform standardelor APA este lucrarea lui Nicol şi Pexman (2010).

19. Din această perspectivă ar fi preferabil ca partea de text să fie mai scurtă – de genul: „Se observă o reducere a simptomatologiei depresive şi o îmbunătăţire a calităţii vieţii la grupul experimental faţă de cel de control, atât la finalul intervenţiei, cât şi după şase luni de la finalizarea intervenţiei (vezi tabelul X)” – decât o descriere încărcată tehnic, care să ofere o informaţie redundantă (lăsând tabelului rolul de a prezenta partea tehnică, adică cifre şi rezultatele testelor cu semnificaţie statistică); această recomandare ar fluidiza şi simplifica partea de prezentare a rezultatelor, eliminând situaţii care prezintă într-o manieră redundantă informaţiile din tabel – ca în cazul unei exprimări de genul: „în privinţa nivelului depresiei se constată că nivelul depresiei la posttest este de 11,25 (2,43) în grupul experimental, faţă de 14,15 (3,15) în grupul de control, această diferenţă fiind semnificativă statistic F (1, 112) = 4,01; p < 0,05 ş.a.m.d.”.

20. Dacă autorii articolului nu apelează la tabele sau figuri pentru a prezenta partea tehnică a rezultatelor, atunci asemenea rezultate pot fi integrate în text după modelul argumentului compus: afirmaţie + dovadă tehnică (exemple: „Angajaţii cu un angajament organizaţional afectiv ridicat manifestă mai puţine comportamente contraproductive: r(54) = –0,39, p < 0,01.”; „Nivelul stimei de sine la adolescenţii care provin din familii monoparentale este

similar cu cel al adolescenţilor care provin din familii normale: pentru cei care provin din familii monoparentale M = 32,12 (AS = 6,54); pentru ceilalţi M = 32,79 (AS = 7,14); t(42) = 0,54, p > 0,10.”).

21. Din aceste tabele nu ar trebui să lipsească indicatorii descriptivi relevanţi precum numărul de participanţi, mediile şi abaterile standard. De asemenea, în cazul studiilor de tip corelaţional supuse unor analize statistice multivariate (regresii, analize factoriale etc.) ar trebui să fie prezentată şi matricea de corelaţii dintre variabile.

22. În cazul în care nu există trimiteri la tabele/figuri, afirmaţiile din text ar trebui susţinute de rezultatele statistice efective (de exemplu, C SMP corelează pozitiv cu indicatorul global C explicit: r(96) = 0,34, p < 0,01). Din motive de fluenţă a textului se preferă însă păstrarea unui text curat, lipsit de elementele tehnice care să îngreuneze citirea şi înţelegerea. În schimb, tabelele (figurile) ar trebui să conţină informaţiile tehnice (dovezile) care susţin aceste afirmaţii.

23. Secţiunea de discuţii conţine adesea referiri la patru dintre cei şase factori identificaţi ca determinanţi în aprecierea calităţii unui articol (Sternberg şi Gordeeva, 1996). Cei patru factori sunt importanţa teoretică, valoarea pentru studiile viitoare, acurateţea metodologică şi importanţa practică a rezultatelor. La aceştia se adaugă doi factori mai generali: actualitatea temei şi calitatea prezentării (adică respectarea recomandărilor precizate în tabelele 8.2-8.6).

24. După cum se poate sesiza pe parcursul întregii cărţi, există o diferenţă între standardele de redactare a bibliografiei conform APA, ediţia a VI-a (APA, 2010), exemplificate în tabelul 8.6, şi standardele adoptate de Editura Polirom pentru colecţiile sale.

25. Dacă articolul nu are alocat un DOI (*digital object identifier*, specificat, de obicei, pe prima pagină a unui articol disponibil online) se trece sintagma „Găsit la” (*Retrieved from*), urmată de adresa electronică a revistei. Dacă revista nu are pagină web (reviste care şi-au încetat apariţia), dar articolul este identificat tot în mediul online, atunci se specifică, în locul adresei web a revistei, adresa web a bazei de date unde a fost găsit (de exemplu, www.eric.com). În sfârşit, dacă articolul există doar în format tipărit, referinţa bibliografică se încheie după precizarea numerelor de pagină, fără a mai fi specificat vreun DOI sau vreo adresă de web.

26. Alte două programe, cu funcţii similare, întâlnite frecvent sunt EndNote şi Reference Manager, însă acestea sunt disponibile contra cost sau sunt dependente de accesul la baza de date ISI Web of Science.

27. În domeniul medical, criteriile grupului Vancouver pentru autorat sunt mult mai dure, solicitând autorilor să îndeplinească concomitent toate cele trei criterii: contribuţie substanţială în conceperea designului şi studiului sau în analiza ori interpretarea datelor; contribuţie substanţială în scrierea sau revizuirea articolului; acordul cu privire la versiunea finală a studiului (vezi Horton, 2008, sau Parmley, 1997).

28. Simmons *et al.* (2011) evidenţiază variaţii în decizia de eliminare a unor participanţi din studii experimentale bazate pe timpi de reacţie (TR), sub motivul că un timp de reacţie prea scurt semnifică o procesare superficială/răspunsuri întâmplătoare. Acest timp prea scurt poate varia

în funcţie de interesele cercetătorului: de la praguri absolute de genul 100 ms, 150 ms sau 200 ms, la exprimări relative în termeni de „x” abateri standard peste medie, unde x poate varia de la un studiu la altul, în funcţie de decizia cercetătorului.

29. Au fost dezvoltate o mulţime de aplicaţii informatice (de exemplu, puteţi căuta pe internet *plagiarism software*), multe dintre acestea fiind disponibile gratuit. Dezavantajul lor este că pot detecta doar pasaje comune scrise în aceeaşi limbă (de exemplu, un articol scris în engleză din surse scrise în aceeaşi limbă), fiind irelevante pentru fraude de tipul unor traduceri extinse (cuvânt cu cuvânt) din engleză în română sau viceversa.

30. Această problemă apare de obicei în domeniul artistic şi în domeniul media, în care creaţii anterioare sunt utilizate pentru a produce efecte/lucrări noi (de exemplu, o Mona Lisa cu mustaţă). Transpunerea acestui principiu în domeniul psihologiei ar însemna permisiunea de a utiliza porţiuni de text scrise anterior în lucrările noi ale aceluiaşi autor sau al altor autori dacă acest fapt are un rol inovator/vizează scopuri diferite. De exemplu, porţiuni dintr-un articol care abordează problematica cost-eficienţă a unei intervenţii psihoterapeutice pot fi reproduse într-un manual introductiv de metodologia analizelor de tip economic în psihoterapie pentru a ilustra modul de realizare a unei astfel de analize, dacă se respectă regulile de citare, respectiv drepturile de autor.

31. Vezi acuzaţiile de plagiat aduse persoanelor din conducerea unei universităţi din România, care apoi au încercat să se disculpe dând vina pe un alt coautor (un doctorand care a abuzat de încrederea lor, plagiind din lucrarea unui alt autor, fără ştiinţa acestora).

32. O listă a revistelor cu acces liber, inclusiv din domeniul psihologiei (158 de reviste) poate fi consultată în DOAJ (*Directory of Open Access Journals* – www.doaj.org). Cele mai multe reviste cu acces liber se autofinanţează prin perceperea unei taxe de publicare de la autorii articolelor acceptate. În ciuda acestui dezavantaj evident, strategia este în continuă extindere datorită celorlalte avantaje majore: scurtarea duratei de publicare, respectiv o probabilitate mai ridicată de acceptare a articolului.

33. Până la urmă, editorul este cel care ia decizia de a publica sau respinge un manuscris. El ia decizii în baza analizei făcute de către evaluatorii desemnaţi şi în baza propriilor reflecţii cu privire la articol. De cele mai multe ori, aceste decizii converg, însă pot exista cazuri în care articolul este acceptat spre publicare în ciuda unor critici majore venite din partea unui evaluator sau situaţii de recomandări favorabile din partea evaluatorilor urmate de decizia editorului de a respinge articolul (însă asemenea situaţii sunt întâlnite doar în cazul revistelor de top, unde concurenţa acerbă împiedică publicarea unor articolele bune).